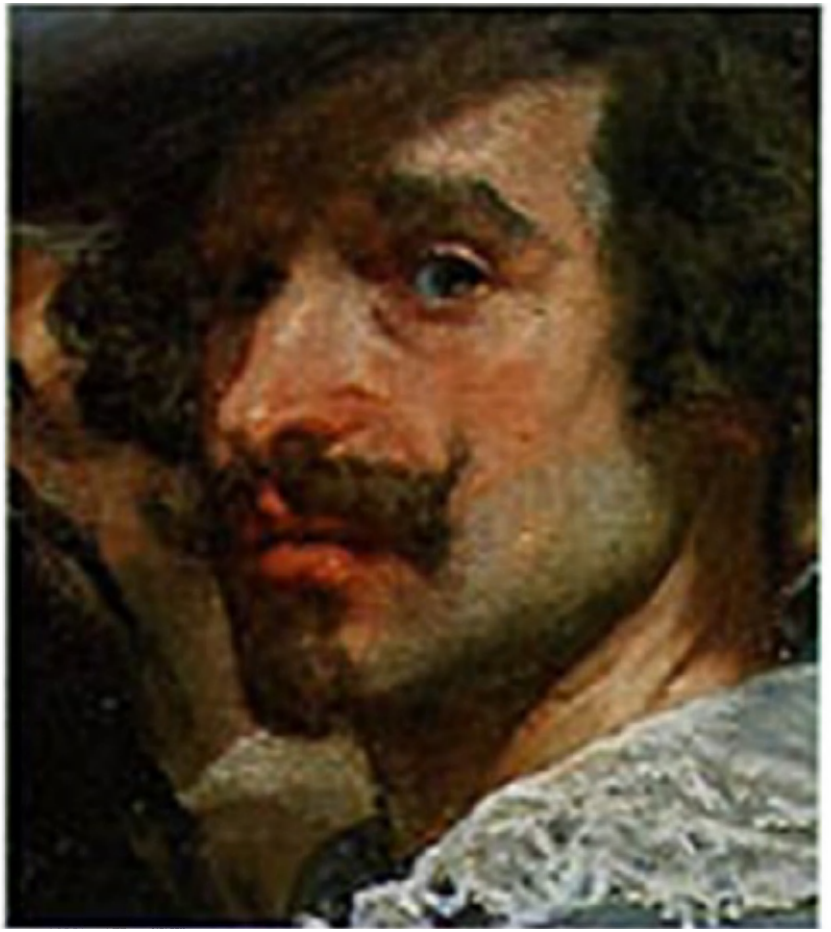
Angel Rivière

Objetos con mente



Alianza
Ps. minor

.lianza Psicología minor

Rivière, A (1991). *Objetos con mente*. Madrid: Alianza.

INDICE

Introducción 13

SECCION PRIMERA: DIFICULTADES PARA EL CONOCIMIENTO DE LA MENTE

Capítulo 1: Obstáculos para un conocimiento objetivo de la mente 21

1.1. El mentalismo objetivo de la psicología cognitiva 21
1.2. Dificultades para el estudio de la mente 25

1.2.1. Tres instantáneas de la historia 25
1.2.2. Escisión y rélación entre las mentes 27
1.2.3. Inferencias sobre inferencias 29
1.2.4. La indeterminación de los modelos cognitivos 31
1.2.5. La conciencia del autómata abstracto: un nuevo fantasma de la máquina 31
1.2.6. ¿Demasiada mente para tan poca? 34

SECCION SEGUNDA: LOS OBJETOS CON MENTE

Capítulo 2: Los objetos con mente y la metáfora del ordenador 39

2.1. La categoría natural de los objetos con mente 39
2.2. El animismo del pensamiento natural y el mecanicismo de la ciencia natural 43
2.3. La categorización de lo mental en la psicología cotidiana 46
2.4. El desafío de Turing 47
2.5. La simplicidad 'de la razón universal 51

Capítulo 3: La máquina abstracta y la mente real: Primeras aportaciones a la ciencia cognitiva 57

3.1. Cálculo de proposiciones en el sistema nervioso: el enfoque de McCulloch y Pitts 57
3.2. Las limitaciones del perceptrón 61

Capítulo 4: Cuatro respuestas al desafío de Turing 67

4.1. Algunas consecuencias de jugar con Turing 67

4.1.1. Dualismo funcionalista 68
4.1.2. El carácter simbólico y discreto de las representaciones 70
4.1.3. Formalismo y sistemas guiados por su sintaxis 72
4.1.4. Flexibilidad mental y sistemas rígidos 73
4.1.5. La irrelevancia computacional de la concencia 74

4.2. Cuatro respuestas al desafío de Turing 76
4.3. La mente de la habitación china y el martirio de la psicología cognitiva 78
4.4. La alternativa literal del Paradigma C-R 87
4.5. El juego de Turing como metáfora: la psicología del procesamiento de la información 93
4.6. Cambiar las reglas del juego de Turing: el enfoque conexionista ... 101

Capítulo 5: Ciencias cognitivas y tecnologías del conocimiento 109

5.1. La(s) ciencia(s) cognitiva(s): ¿Singular o plural? 109
5.2. Un enfoque nuevo de viejos problemas epistemológicos. Psicología cognitiva y tecnologías del conocimiento 115

SECCION TERCERA: LAS FUNCIONES MENTALES

Capítulo 6: Psicología natural y enfoques cognitivos de las funciones mentales 129

6.1. Conocimiento y comprensión natural de la mente 129
6.2. La defensa cognitiva de las actitudes proposicionales 135

Capítulo 7: La atribución de lo mental. Consideraciones sobre la psicología natural 147

7.1. El lenguaje de lo mental: un diálogo con Archer 147

7.2. Los primates mentalistas y las ventajas evolutivas del engaño y su reconocimiento 154
7.3. Psicología natural, intensionalidad y metarrepresentación 161
7.4. La conciencia como sistema de predicción mentalista y la semántica de los verbos mentales 163
7.5. Psicología natural y psicología popular 166

Capítulo 8: Los enunciados intencionales y la búsqueda de objetividad en psicología 171

8.1. Mentalismo, animismo y evolución de la razón 171
8.2. Funciones mentales y objetividad de las explicaciones científicas.... 173
8.3. Los enunciados con verbos mentales como obstáculos epistemológicos 176
8.4. Las alternativas de tratamiento científico de los enunciados con verbos mentales 178

Capítulo 9: El pensamiento y el lenguaje como funciones mentales. Algunas reflexiones sobre intencionalidad, intensionalidad y forma 185

9.1. Pensamiento y lenguaje: algunos enfoques de la "psicología popular" 185
9.2. Pensamiento, lenguaje e intensionalidad 187
9.3. La intuición de la forma en el pensamiento y el lenguaje 189
9.4. La intuición de la forma y la psicología cognitiva 192

Capítulo 10: Cuatro enfoques psicológicos de las funciones mentales 195

10.1. Introspección, objetividad y sistemas de cómputo 195
10.2. Funciones mentales y psicología introspectiva: consideraciones históricas 198
10.3. La reducción extensional del conductismo 203
10.4. El origen del paradigma C-R: una mecánica de la intencionalidad 207
10.5. El conexionismo y la recuperación de los objetos extensionales como conceptos teóricos en psicología 220

Bibliografía 229

INTRODUCCION

En un experimento muy reciente, en que se empleaban como estímulos reproducciones de rostros pintados por Velázquez como los que figuran en la portada de este libro, el doctor Simon Baron-Cohen y el autor de estas páginas encontraban un resultado que era esperable, por lo que sabemos intuitivamente de las personas, pero que merece una consideración detenida cuando se somete a una reflexión algo más profunda: a pesar de los más de tres siglos de distancia y de las enormes diferencias culturales que nos separan de aquel pintor genial, los sujetos del experimento —ingleses y españoles— eran capaces de distinguir, de forma inequívoca, las **experiencias emocionales,** tales como la ira, la desconfianza, el disgusto, etc, reflejadas en los rostros pintados por Velázquez.

Esa capacidad de **leer la mente** es, sin duda, uno de los rasgos más acentuados y profundos de los miembros de nuestra especie. Los humanos somos individuos mentalistas, psicólogos naturales notables, que no sólo interpretamos emociones (y, más aún, las revivimos) cuando las vemos expresadas por otros, sino que constantemente inferimos las creencias y deseos de los demás, razonamos sobre sus estados mentales, y empleamos un lenguaje mentalista de una riqueza considerable, compuesto de un vocabulario de verbos, tales como «pensar», «intuir», «saber», «descubrir», «suponer», etc, y de nombres tales como «idea», «pensamiento», etc. Basta con detenerse a escuchar cualquier conversación normal (como haremos en el capítulo séptimo de este libro), o con pararse a observar cualquier interacción cotidiana entre humanos, para caer en la cuenta de hasta qué punto están moduladas nuestras relaciones interpersonales por el pensamiento y el lenguaje sobre lo mental.

Esta observación es tan cotidiana y evidente que parece una de esas trivialidades inevitables de nuestra vida normal de las que no cabe esperar consecuencias conceptuales importantes. Sin embargo, en este caso como en tantos otros, puede ser una estrategia útil del científico la de desconfiar de lo evidente, buscando significados profundos más allá de lo que parece trivial. Al fin y al cabo, los fenómenos de caída de los cuerpos se observan centenares de veces por todo ser humano y, sin embargo, resultaron ser fenómenos interesantes en sí mismos cuando fueron analizados desde la perspectiva de la **nuova scienza** física que nació en el renacimiento. ¿Qué sucede cuando el psicólogo, como científico, analiza los fenómenos, que se dan diariamente, en los que atribuimos,, de forma implícita, una **mente** a ciertos objetos, al decir de ellos que piensan, creen, perciben, desean, etc.?

¿Qué significa atribuir mente a ciertos objetos, sean éstos personas, animales, nubes siderales (como en una novela de Stanislav Lem), seres divinos, fenómenos naturales o artefactos tales como los ordenadores digitales?, ¿de qué objetos se predica que tienen mente? Estas cuestiones son, en sí mismas, interesantes para el psicólogo, que al fin y al cabo es un «profesional de la mente». Pero sucede que, además de ser fascinantes por sí mismas, las respuestas a tales preguntas pueden proporcionarnos herramientas conceptuales de una gran utilidad para el análisis crítico de la psicología científica. Cuando contraponemos el lenguaje sobre la mente de la «psicología natural» (que empleamos cotidianamente como instrumento para comprender y predecir la conducta de los otros) con el lenguaje y las observaciones acerca de lo mental que provienen de la psicología científica, obtenemos una perspectiva de análisis de ésta que nos permite entender mejor las raíces conceptuales de la ciencia de la mente.

Éstos son los propósitos que se plantea este libro (y, al decirlo así, asumimos que estamos empleando significativamente un lenguaje mentalista y animista plenamente admitido; los libros «no se plantean propósitos»): se trata de realizar un análisis crítico de la psicología científica —y en especial de la parte de ella conocida como «psicología cognitiva»— mediante el uso de herramientas conceptuales que provienen del análisis de la llamada «psicología natural», es decir, la que usamos de forma cotidiana e implícita en nuestras relaciones habituales con otros humanos.

Tal análisis crítico de la psicología cognitiva parte de la premisa, analizada en el capítulo primero, de que es difícil hacer una ciencia objetiva de la mente: y lo es, antes que nada, porque el concepto de «mente» tiene, cuando menos, dos significados diferentes en psicología. Se identifica, por una parte, con la conciencia. Por otra, con un sistema de conocimiento de naturaleza computacional. ¿Cómo puede la psicología conocer ese sistema y qué dificultades encuentra para ello?, ¿cómo puede hacerse consciente, por otra parte, un sistema de cómputo? No creo que nuestra reflexión sobre los objetos con mente proporcione, ni mucho menos, una respuesta completa a estas preguntas. Espero, por lo menos, que tenga la virtud de haberlas formulado de forma correcta y suficientemente profunda como para favorecer la reflexión del lector.

Si tiene, el lector, esa mínima paciencia que exige llegar al capítulo primero, obtendrá la ventaja de inquietarse aún más de lo que pueda estarlo con esta introducción, pues en ese capítulo se examinan los obstáculos que tenemos los psicólogos para desarrollar un conocimiento objetivo acerca de lo mental. Espero que el capítulo primero, como en las buenas novelas policiacas de Ross MacDonald (del que hablaremos algo más en otro capítulo del libro), tenga la virtud de intrigar al paciente lector lo suficiente como para dar un paso más.

Lo que encontrará el lector en el capítulo segundo es una reflexión acerca de los objetos con mente, que parte de una premisa muy simple y espero que universalmente aceptable. La premisa es ésta: las personas dividimos nuestro mundo de objetos en dos grandes grupos. De unos decimos que tienen mente, mientras que a otros no les atribuimos esta propiedad. Esa sencilla reflexión nos llevará a enfrentarnos al problema, inaugural de la psicología cognitiva, del significado de atribuir una función prototípicamente mental, tal como la de «pensar», a los ordenadores. En los capítulos tercero y cuarto se examinan las consecuencias e implicaciones de la llamada «metáfora del ordenador» en psicología, y las diferentes respuestas dadas a la pregunta sobre si los ordenadores piensan, y a la cuestión complementaria de si las mentes son, en realidad o en metáfora, sistemas de cómputo. En el capítulo quinto se examinan las profundas relaciones que existen entre la psicología y las otras ciencias cognitivas y tecnologías del conocimiento. Este conjunto de capítulos constituye la sección segunda del libro, que se titula «Los objetos con mente».

La sección tercera plantea un problema que es complementario con el anterior: prescindiendo ahora de cúales sean los objetos a los que atribuimos la propiedad de tener funciones mentales, ¿qué son esas funciones? El capítulo sexto da una primera respuesta a esta pregunta. En

el séptimo se reflexiona sobre el hecho de que los humanos (y quizá otros primates) seamos seres capaces de suponer que los demás tienen funciones mentales y de acceder a la conciencia de que nosotros mismos las tenemos. Finalmente, los tres capítulos finales analizan el lenguaje mental de la psicología científica, partiendo de una observación un tanto intrigante: la de que el lenguaje y el pensamiento mentalistas han sido siempre, simultáneamente, puntos de partida de la psicología como ciencia y serios obstáculos epistemológicos para su constitución como tal. En el último capítulo se examinan cuatro alternativas básicas en el enfrentamiento a las funciones mentales: la primera (psicología introspectiva) implica el empleo de un lenguaje mental tanto en los enunciados observacionales de la psicología como en los teóricos. La segunda (conductismo) exige la eliminación de! vocabulario mentalista de unos y otros enunciados. La tercera (psicología cognitiva clásica) emplea un lenguaje observacional no mentalista, al que da significado con un lenguaje teórico mentalista. La última y más reciente (conexionismo) permite comenzar a comprender cómo puede surgir la enigmática propiedad de tener mente de objetos materiales tales como las redes neurales del cerebro.

Espero que el hecho de haber «anticipado el final» (algo que Ross MacDonald nunca se hubiera permitido) no desanime al lector de pasar más allá de esta breve introducción. El tiene la última palabra con respecto a la justificación de los argumentos que se proponen, a lo largo de nuestra reflexión sobre los objetos con mente, en esa contraposición entre la psicología cognitiva y la psicología natural. El autor cree que la una y la otra se comprenden mejor cuando se examinan sus estrechas, volubles y sorprendentes relaciones. Cree también que es tan útil el análisis de la psicología cognitiva con los conceptos que provienen de la psicología natural, como el inverso: el examen de la psicología natural con esquemas cognitivos.

Creo que las breves consideraciones anteriores son suficientes para que el lector pueda delimitar con claridad el carácter de este libro y de lo que puede encontrarse en él: no hallará una «introducción a la psicología (cognitiva)», ni —menos aún— un «manual de psicología». Sí un ensayo reflexivo sobre los fundamentos de la psicología cognitiva que, si es que tiene alguna virtud, ésta será al menos la de una cierta originalidad: un tipo de análisis que, por lo que sabe el autor, no se había realizado anteriormente con el mismo tipo de conceptos que aquí se utilizan. El autor espera que no pueda aplicarse, sin embargo, a este ensayo el mismo comentario terrible que hacía un escritor genial (Jorge Luis Borges) a un autor novel: «Su novela tiene partes buenas y partes originales. Las originales no son buenas. Las buenas no son originales».

En cualquier caso, de las partes no buenas de esta obra no son responsables algunas personas que han colaborado, de forma eficaz y frecuentemente abnegada, a su elaboración. Los investigadores del grupo de «Teoría de la Mente» de la Universidad Autónoma de Madrid, Encarnación Sarriá, María Nuñez, María Sotillos, Juan Rubio, Laura Quintanilla, Lina Arias, han tenido la paciencia de leer cuidadosamente —y criticar perspicazmente— el manuscrito, antes de su edición. También tuvieron antes la paciencia de soportar al autor en el periodo de escritura del libro.

Este último mérito es aún más destacable en personas que convivieron más cercanamente con el autor, como Inés Marichalar (que además dedicó su tiempo a cuestiones tan poco gratificantes como las relacionadas con la ordenación bibliográfica), Mario, Lucía, Inés y Pablo Rivière. Por otra parte, Jaime Riviére dedicó más noches de las debidas a ayudar en el tratamiento informático final'del texto y hacer comentarios inteligentes sobre distintos aspectos de su contenido.

Finalmente, varios compañeros de la Universidad Autómoma de Madrid han colaborado, de forma eficaz y amistosa, a que este libro pudiera llevarse a cabo. Durante el periodo de su elaboración fueron de enorme valor los comentarios y críticas de Mario Carretero. En la fase última de escritura, Juan Antonio Huertas y Mercedes Belinchón suplieron algunas ausencias de «la vida real» del autor.

La terminación del libro se ha visto muy facilitada por la concesión de la ayuda a la investigación PB-89-0162, del programa sectorial de promoción del conocimiento, por parte de la Dirección General de Investigación Científica y Técnica (DGICYT) del Ministerio de Educación y Ciencia. Esta ayuda dio, en primer lugar, ánimos y, en segundo, recursos al autor para realizar un conjunto de investigaciones sobre el pensamiento acerca de lo mental, estrechamente relacionadas con el contenido del libro.

Sección primera:
DIFICULTADES PARA EL CONOCIMIENTO DE LA MENTE

Capítulo 1:
OBSTACULOS PARA UN CONOCIMIENTO OBJETIVO DE LA MENTE

1.1. El mentalismo objetivo de la psicología cognitiva

En la segunda mitad de nuestro siglo, los psicólogos hemos desarrollado un nuevo vocabulario teórico, y procedimientos empíricos originales, para volver a nuestro viejo propósito de estudiar la mente. El vocabulario incluye términos como «esquema», «proposición» o «cómputo». Los procedimientos, actividades como medir y analizar latencias en la comprensión de oraciones, investigar los errores que cometen las personas cuando razonan, o registrar sus pequeños movimientos de ojos mientras están leyendo un texto. Se han propuesto nombres diversos para el conjunto formado por las palabras y reglas del lenguaje teórico, las operaciones prototípicas de investigación empírica y los presupuestos de ambas cosas: *paradigma cognitivo, psicología del procesamiento de la información,* etc. Todos estos nombres son discutibles, con argumentos tales como que no es adecuado el término «paradigma» para los nuevos modelos de la mente (Caparros y Cabucio, 1986), ni es estrictamente «información», sino conocimiento, lo que se suele definir en los modelos cognitivos (De Vega, 1989; Dretske, 1981). Dichas críticas —en las que no debemos detenernos ahora— son bastante razonables, por lo que emplearé aquí las dos palabras más neutras, y más frecuentes, con las que se etiqueta el «nuevo» enfoque: *psicología cognitiva.*

Me propongo, en estas páginas, razonar sobre los fundamentos de la psicología cognitiva. Para ello me serviré de un concepto, el de *mente,* que se nos presenta como intuitivamente obvio en nuestras interacciones habituales con los demás y con nosotros mismos, pero que ha demostrado ser, a lo largo de la historia, mucho más complejo y huidizo de lo

que parece a primera vista. ¿Por qué seleccionar, entonces, un concepto tan impreciso, tan esquivo, tan rebelde a cualquier intento de sometimiento al pensamiento riguroso?

Hay varios argumentos a favor de esta opción: en primer lugar, el concepto de mente, precisamente por su liberal amplitud, deja sitio holgado donde poner nuestras razones. Además tiene una vieja tradición psicológica, que permite encadenar esas razones a la historia, evitando que caigan en el vacío de lo intemporal. Finalmente, está el hecho de que, por muchos desacuerdos fundamentales que puedan existir sobre el objeto de la psicología cognitiva —¿los procesos mentales son parecidos a las computaciones que se realizan con ordenadores? Si lo son, ¿sobre qué medio se realizan las computaciones?, ¿son representaciones simbólicas lo que efectivamente se computa?—, todo el mundo está de acuerdo en que trata de la mente.

Un mínimo común denominador en que todos los psicólogos estaríamos conformes es ciertamente éste: *la psicología cognitiva pretende ser una ciencia objetiva de la mente, a la que concibe como sistema de conocimiento.* Creo que no habría desacuerdos en reconocer ese intento: A diferencia de la vieja psicología introspectiva de comienzos de este siglo, y del conductismo que dominó hasta su segunda mitad,- la psicología cognitiva (a) trata con la mente, (b) ha abandonado, en gran parte, Ta vieja ilusión de que es posible hacer ciencia a base de «hablar directamente con ella» sin intermediarios, por acceso introspectivo; (c) por- ello, se ve obligada a reconstruir indirectamente la estructura y los procesos de la mente, sirviéndose de procedimientos objetivos, y (d) la trata, sobre todo, en tanto que es un sistema capaz de conocer.

Desde ahí, todo serán desacuerdos entre los psicólogos, siempre fieles a nuestras inveteradas propensiones polémicas y ricas tradiciones de pluralidad paradigmática: Independientemente de sus pretensiones, ¿consigue la psicología cognitiva hablar realmente de la mente, o se deja algo esencial de ella en el intento?, ¿lo hace, como supone, de forma objetiva o sigue siendo una psicología peligrosamente subjetivista?, ¿logra explicar el conocimiento, o lo reduce a unos términos tales que hacen imposible comprender su origen, naturaleza real y significado?

Dado que tenemos el objetivo de razonar y justificar esas pretensiones, no parece lógico que demos por sentada una respuesta positiva a las tres preguntas anteriores. Muy al contrario, nuestras reflexiones sólo serán rigurosas si parten de un *mínimo* de presuposiciones, y si hacen que éstas sean tan evidentes como sea posible. En eso, deberíamos acercarnos lo más posible a las máximas para un pensamiento preciso que proponía un precursor, Renato Descartes, que nos ha enseñado a los psicólogos cognitivos otras muchas cosas además de éstas:

> Esas largas cadenas de razones, completamente simples y fáciles, de que los geómetras suelen servirse para llegar a sus demostraciones más difíciles, me habían dado ocasión de pensar que todas las cosas que pueden caer bajo el conocimiento de los hombres, se encadenan de igual forma, y que, con tal de abstenerme de admitir por verdadera una que no lo sea y de mantener siempre el orden preciso para deducirlas unas de otras, no puede haber ninguna, por más lejos que se hallen situadas, a las que no se llegue en última instancia, ni tan ocultas que no se descubran. (*Discurso del Método,* edición de 1989, p. 35).

Seguiremos, entonces, las buenas costumbres de los geómetras, que suelen usar un procedimiento muy útil cuando hilvanan sus razones: decir, antes que nada, de dónde parten y a dónde quieren llegar. Veamos primero a dónde queremos ir a parar. ¿Cuáles son nuestras tesis finales? En resumen, dicen lo siguiente:

1. Es posible una ciencia que sea, al mismo tiempo, objetiva y mentalista. Mentalista aquí no quiere decir sólo que trate acerca de la mente, sino que emplea un vocabulario que contiene *términos intencionales, intensionales y epistémicos* para hablar rigurosamente acerca de ella. Términos, por tanto mentales y, más en concreto, términos que se refieren a *representaciones mentales.*

2. La posibilidad de tal ciencia se ha basado históricamente en la metáfora computacional: en la idea de que la mente puede comprenderse mejor si se entiende como un sistema de cómputo semejante, en cierta medida, a un ordenador digital. Esta metáfora ha tenido una gran importancia en la justificación de la posibilidad de una ciencia objetiva de la mente como sistema de conocimiento. La razón de esa importancia es que, por su parte, la metáfora presupone que pueden definirse autómatas capaces de dar cuenta, de forma universal, de los procesos cognitivos. De dar explicación de ellos sin ninguna clase de conceptos circulares o pseudo-explicativos, como los que denunciaba Ryle (1949) cuando criticaba a todas las psicologías mentalistas, diciendo que éstas partirían implícitamente de un supuesto dudoso: el de la mente como «un fantasma en una máquina», «the ghost in the machine». Así, la psicología cognitiva se define por un enfoque de *mecanicismo* abstracto.

3. Por consiguiente, un aspecto esencial de la ciencia de la que hablamos es que entiende las funciones de conocimiento como procesos de

UNIVERSIDAD DE ANTIOQUIA
BIBLIOTECA CENTRAL

cómputo sobre representaciones, sobre la base de los supuestos del enfoque del «mecanicismo abstracto» al que me he referido. Esta característica se manifiesta con mayor claridad en el enfoque aún dominante en psicología cognitiva. Existe, sin embargo, una sensibilidad creciente hacia las anomalías que presenta ese paradigma dominante: no logra explicar el papel de la conciencia en el conocimiento, tiende a concebirlo en términos excesivamente logicistas y lo descontextúa de sus condiciones biológicas y sociales.

4. El paradigma dominante, al que llamaremos C-R (de las computaciones sobre representaciones), ha sido cuestionado, en los últimos años, desde posiciones que o bien (a), al tener en cuenta ciertas propiedades del sistema nervioso, conciben' los procesos cognitivos como cómputos en paralelo que se realizan en un sistema de unidades que representan, de forma colectiva y distribuida, información de carácter no-simbólico (conexionismo), o (b) al considerar el carácter social, intencional y creativo del conocimiento, muestran un disgusto creciente con relación al mecanicismo abstracto y el solipsismo del modelo C-R clásico (enfoque de la «enacción», tendencias interaccionistas y propensiones neopragmatistas no bien definidas, etc).

5. Así, no hay una sola *psicología cognitiva* (en sentido amplio) sino varias. El conocimiento puede explicarse psicológicamente de varias maneras. Todas ellas comparten algunos supuestos mínimos: (1) hay un nivel, por lo menos relativamente autónomo, de explicación de la mente, en que el funcionamiento de ésta se entiende condicionado por procesos y representaciones de conocimiento, (2) la investigación de tales procesos y representaciones no depende del acceso introspectivo. Términos tales como «operación» (Piaget, 1947), «esquema» (Bartlett, 1932; Bransford y Johnson, 1973; Rumelhart, 1980; Rumelhart, Smolensky, McClelland y Hinton, 1986), «proposición» (Pylyshyn, 1984), «plan» (Miller, Galanter y Pribram, 1960), «imagen mental» (Kosslyn, 1980), que se refieren a procesos o estructuras *que se dan en la mente*, no se justifican por su evidencia fenoménica a la introspección sino por su valor para explicar y predecir la conducta. Forman parte del vocabulario de la psicología cognitiva en sentido amplio.

6. Las formas en que se organizan las estructuras y procesos a que se refieren estos términos se *reconstruyen* a partir de datos objetivos. Así llegamos a nuestro punto de partida: en su significado más general, *la psicología cognitiva implica un enfoque mentalista y objetivo de los proceso psicológicos y, en especial de aquellos que permiten conocer.*

Ahora podemos volver a nuestro objetivo primero: se trata, decía mos, de justificar y dar fundamento a las afirmaciones que aparecen en los seis puntos anteriores y que, como acabamos de ver, pueden reducirse a una fundamental: es posible una psicología mentalista y objetiva del conocimiento. Ahí queremos ir. ¿Cuáles pueden ser nuestros puntos de partida? Tenemos que buscar, como hacía Descartes y hacen los geometras, algunas —si es posible pocas— razones ciertas y verdades evidentes para dar origen a nuestra cadena de razonamiento. Pueden ser por ejemplo, estas dos: (1) Conocer objetivamente la mente es difícil (2) Las personas tendemos a dividir el mundo en dos grandes categorías de objetos: de unos decimos directa o indirectamente que tienen mente. De otros no lo decimos.

1.2. **Dificultades para el estudio de ia mente**

1.2.1. *Tres instantáneas en la historia*

Es difícil, sin duda, hacer una ciencia objetiva de la mente. Una buena prueba de ello es que los psicólogos hemos desmayado, más de una vez, en nuestro viejo y tenaz esfuerzo de hacerla. Veamos rápidamente una pequeña ilustración histórica de esos esfuerzos y desmayos.

Hace justamente un siglo, se publicaba uno de los libros más reflexivos y brillantes de toda la historia de la psicología, *The Principies of Psychology* de William James (hay una edición en español de 1990). Como no debemos detenernos mucho en esta ilustración, haremos lo que los malos lectores de novelas policíacas: leer solo la primera y la última página. La primera empieza así: «La Psicología es la Ciencia de la Vida Mental, tanto de sus fenómenos como de sus condiciones. Los fenómenos son cosas del tipo de aquellas a las que llamamos sentimientos, deseos, cogniciones, razonamientos, decisiones y cosas por el estilo» (p. 15. Edición inglesa de 1983). La última, de forma realmente sombría y desmayada, termina así: «Aun en las porciones más claras de la Psicolog a nuestra penetración es insignificante. Y cuanto más sinceramente nos esforcemos por trazar el curso actual de la *psicogénesis,* es decir, los pasos por medio de los cuáles hemos ido adquiriendo los atributos mentales peculiares que poseemos como especie, más claramente percibiremos "el lento avance del crepúsculo hacia el seno de la noche"» (ibíd. p. 1100).

¿Era sólo el esfuerzo de escritura lo que le tenía tan pesimista a James?; ¿estaba aquel día de peor humor de lo normal? No, no es ésa la razon; era, más bien, su constatación de los logros «insignificantes» de una psicología introspectiva sobre los *fenómenos* de la mente la que le llevaba al desánimo. Una psicología que estaba compendiada y revisada cuidadosamente en los *Principles.*

Demos ahora un salto de cuarenta años. Estamos en 1930. Watson es, en *Behaviorism*, mucho más optimista que James. Pero lo es, sencillamente, porque ha *renunciado* a esa tarea difícil de conocer la mente. Cuando habla, por ejemplo, del pensamiento dice lo siguiente: «Antes de comprender la teoría conductista del pensar, ¿no querría el lector abrir cualquier tratado de psicología introspectiva y leer el capítulo que se le dedica? ¿No querría tratar de digerir algo del pábulo brindado por los filósofos acerca de esta importantísima función? Nosotros hemos intentado hacerlo y tuvimos que renunciar. Creemos que el lector también renunciaría/../ Tanto cuando nos sentábamos en el regazo de nuestra madre como en los laboratorios psicológicos, se nos enseñó a decir que el pensamiento es algo particularmente incorpóreo, algo sumamente intangible, sumamente evanescente, *algo especialmente mental /../* En razón de la naturaleza oculta de la musculatura que lo realiza, el pensar siempre ha sido inaccesible a la observación y a la experimentación directa /../ A medida que se descubren nuevos hechos científicos, se reduce cada vez más el número de fenómenos inaccesibles a la observación y, por consiguiente, también el número de "perchas para colgar el misticismo"» (pp. 225-226. Ed. esp. de 1972).

Ya no se percibe aquí «el lento avance del crepúsculo hacia el seno de la noche», sino el amanecer radiante del conductismo, que ha tenido, sin embargo, un precio: la renuncia a cualquier vocabulario mental. La consideración de lo mental como «una percha donde colgar el misticismo».

Más de medio siglo después, un psicólogo cognitivo, Philip Johnson-Laird (1988) vuelve a sentirse optimista, sin dejar de ser mentalista. Su comentario no tiene nada ni de la sombría desesperanza del de James, ni del confiado reduccionismo de lo mental a movimiento muscular que veíamos en Watson. Nos dice Johnson-Laird lo siguiente: «La invención del ordenador digital programable, y de manera más importante, su precursora, la teoría matemática de la computabilidad, ha obligado a la gente a pensar de una forma nueva sobre la mente. Antes de la computación había una distinción clara entre cerebro y mente; uno era un

órgano físico y la otra una «no entidad» fantasmática que difícilmente resultaba un tema de investigación respetable. (Se consentía que los adultos pudiesen hablar de ella en privado, siempre y cuando comprendieran que, en realidad, no existía.) Después de la llegada de los ordenadores no cabe semejante escepticismo: una máquina puede controlarse mediante un «programa» de instrucciones simbólicas, y no hay nada de fantasmal en un programa de ordenador. Quizá, y en gran medida, la mente es para el cerebro lo que el programa es para el ordenador. De esta manera, *puede haber una ciencia de la mente»* (pp. 13-14. Ed. esp. de 1990, el subrayado es nuestro).

Son instantáneas muy rápidas de tres momentos de las actitudes psicológicas hacia la mente: (1) la psicología es la ciencia de los *fenómenos* de la mente, dice James, a los que accede por introspección. James (como también Wundt; vid. Danziger, 1980) es poco optimista sobre las posibilidades de progreso científico por ese camino, pero no concibe otro (Rivière, 1990); (2) la psicología *no* es la ciencia de la mente, para Watson. Hay que abandonar cualquier vocabulario mentalista y hablar sólo de observables; (3) puede haber una ciencia de la mente, objetiva, y basada en el supuesto de que la mente es un sistema de cómputo. ¿Resuelve esta tercera respuesta las interrogantes planteadas por las otras dos? Da solución a muchas, pero no a todas. Veamos un ejemplo.

1.2.2. *Escisión y relación entre las mentes*

La posición de James se basaba en una fusión entre mente y conciencia que, como veremos en otro momento, ha tenido que ser rota por la psicología cognitiva (como antes lo fue, en otras direcciones, por Leibniz, Herbart, Helmholtz, Freud, etc). Aunque no se suela destacar, un supuesto esencial de ésta es que el aparato consciente al que tenemos acceso introspectivo, esa mente de los *fenómenos* de que hablaba William James (1890), no se identifica con el sistema que computa representaciones simbólicas al que se refiere Johnson-Laird (1988). Para el lector de estas páginas, esta afirmación puede ser intuitivamente comprensible de forma muy sencilla: lo que tiene en la cabeza, si presta atención al texto, es seguramente un conjunto de contenidos conscientes que se refieren a las palabras que lee, las oraciones que comprende, las posibles intenciones y significados del texto. Sin embargo, al lector le son completamente inaccesibles los procesos mentales que, por debajo

de su conciencia y más allá de sus fronteras, le permiten realizar actividades tales como acceder al sistema léxico representado en su memoria, inferir proposiciones del texto, analizar implícitamente algunos de los aspectos de la organización sintáctica de las oraciones, etc. Estos son, seguramente, procesos de *cómputo*, que realiza la mente del lector con gran eficacia y sin contar apenas con atisbos conscientes. Lo que llega a la conciencia es sólo una pequeñísima parte de los resultados finales o intermedios de los procesos computacionales subyacentes.

La mente fenoménica *no* se identifica, en efecto, con la mente de los cómputos. Por ejemplo, la mente fenoménica tiene un lenguaje (el habla interna, las imágenes mentales, etc.) y la computacional otro diferente (quizá, las proposiciones, estructuras conceptuales, etc). Ha costado un largo siglo de investigación psicológica pasar de la que podemos llamar «mente uno» —la de James— a la «mente dos» —la de Johnson-Laird—. Un siglo de avances decisivos, pero que siguen dejando en el aire cuestiones que no son menores, sino realmente sustantivas. Por ejemplo, ésta: ¿cuál es la relación entre la mente-uno y la dos?

Ray Jackendoff (1987) ha hecho un brillante análisis de este tema, al que debemos referirnos ahora brevemente: señala que, del mismo modo que existía un problema de relaciones entre la mente fenoménica (aquella a la que llamábamos «uno») y el cuerpo, así también lo hay para relacionar la mente fenoménica y la computacional. Y, de la misma manera que ante aquel problema de relaciones cabían alternativas diferentes, tales como el interaccionismo (un estado cerebral, $e.c_1$, produce un estado mental, $e.m_1$ y éste es *causa* de otro, $e.m_2$, que a su vez causa un nuevo $e.c_2$), el epifenomenalismo (sólo hay causas en el dominio de los estados cerebrales. Los otros son epifenómenos), el paralelismo (estados cerebrales y estados mentales se corresponden, pero no interactúan causalmente) y la teoría de la identidad (estados cerebrales y mentales son la misma realidad, descrita de diferentes formas), también hay opciones distintas para enfrentarse al problema de la relación entre «mente uno» y «mente dos». ¿Es la conciencia un puro epifenómeno de las computaciones?, ¿son realidades paralelas que no se influyen? ¿es posible que interactúen? Y, si lo es, ¿cómo?. ¿No serán sólo dos formas de hablar de lo mismo?

La realidad es que antes teníamos un solo problema (el tradicional, e incluso entrañable, problema mente-cuerpo) y, ahora, dos: tenemos ése y además un problema $mente_1$-$mente_2$ (Jackendoff, op. cit.). Hemos conseguido la magia de convertir un problema en dos. Como en ese

divertido poema de Angel González, que se titula *Ciencia aflicción:* «Si todo problema al resolverse plantea más problemas/ dentro de poco será difícil andar por las calles. /—*Guau, guau,!* nos dirán los problemas enseñándonos los dientes,/ mordiéndonos los fondillos de los pantalones,/ aturdiéndonos con sus bufonadas insolubles./ Si todo problema/ —como viene sucediendo hasta ahora—/ plantea dos problemas,/.../».

Lo paradójico de la situación actual es que, mientras parecen vislumbrarse vías de solución al problema de las relaciones entre el organismo y la mente computacional, a partir de la consideración conexionista de modelos abstractos del sistema nervioso como sistemas de cómputo (McClelland, Rumelhart y Grupo de Investigación PDP, 1986), las relaciones entre la mente fenoménica y la computacional siguen siendo muy oscuras, y la falta de explicación de ellas una de las anomalías mayores de la psicología cognitiva (vid., sin embargo, Baars, 1988; Johnson-Laird, 1988; Jackendoff, 1987): el viejo problema mente-cuerpo empieza a ser, por lo menos tratable, pero no lo es tanto el nuevo problema $mente_1$ -$mente_2$ que se ha suscitado a partir de la concepción del sistema cognitivo como un mecanismo de cómputo. Así, a pesar de los progresos importantes en el estudio de la mente que abonaban el optimismo de Johnson-Laird (op. cit.), no parece que debamos prescindir, por ahora, de la humilde y trivial observación que habíamos tomado como premisa de nuestra reflexión: «conocer objetivamente la mente es difícil».

1.2.3. *Inferencias sobre inferencias*

Hay además otros factores que influyen en esa dificultad: Uno de ellos se deduce, de manera evidente, de lo que acabamos de decir. Si la mente que efectúa los cómputos —la mente que emplea esquemas, posee conceptos, comprende proposiciones, etc— *no* puede identificarse descriptivamente con aquella otra que presenta a la mirada de la conciencia *fenómenos* tales como las imágenes mentales y las palabras que nos decimos «en silencio» a nosotros mismos (aunque pocos duden que una y otra corresponden a una misma realidad), eso significa que no hay ninguna constancia fenoménica directa de la mente computacional. Esta tiene que ser *reconstruida conceptualmente* a partir de la conducta de la gente en situaciones generalmente muy controladas. Gillian Cohén explica muy bien el problema en el capítulo sobre memoria semántica de su manual de *Psicología cognitiva* (1977):

Obviamente, una de las mayores dificultades para la investigación experimental de las operaciones cognitivas es la inaccesibilidad del fenómeno que se estudia-. Los trabajos de la memoria semántica tienen lugar en el circuito más interno de una cadena de procesamiento de la cual solo dos terminales, el estímulo o entrada final y la respuesta o salida final son observables. En consecuencia, las conclusiones experimentales son necesariamente inferidas y los resultados de inr cluso los experimentos mejor diseñados necesitan ser interpretados. El estudio de la memoria semántica es más bien un intento de averiguar cómo están clasificados, ordenados y catalogados los libros de una gran biblioteca de préstamo sin permitir la entrada al investigador. Supongamos que el investigador sólo puede sentarse a la puerta de entrada y preguntar a los lectores que toman prestados ios libros qué es lo que están buscando según van entrando, y anotar la duración de la visita y el número y tipo de libros que sacan. Es dudoso que la información que tiene a su alcance sea la suficiente como para darle una idea exacta de lá disposición y el sistema con que está organizada la biblioteca, aunque podría ser capaz de formular algunas conjeturas con fundamento. En el estudio de la memoria semántica, incluso las técnicas experimentales más rigurosas e ingeniosas se encuentran limitadas de forma similar (op. cit., p.p. 2-3 de ed. esp. de 1983).

De la mente fenoménica hay una «percepción interna»: una *innere Wahrnehmung,* decía Wundt (1887), en la que se basa la posibilidad de acceder introspectivamente a ella. Esa percepción interna es previa a la auto-observación de la mente (*Selbstbeobachtung)* y no debe confundirse con ésta. Es la misma distinción que hacía Brentano (1874), que añadía además la afirmación de que la *innere Wahrnehmung* proporciona una evidencia que no es comparable con la que brinda la percepción externa de los objetos del mundo. Una evidencia más cierta y, como se ha dicho, «incorregible». Una evidencia de *fenómenos,* que tienen lugar en el espacio interno (Brentano) o inmediato (Wundt) de la experiencia. Sin embargo, de la mente como sistema de cómputo no hay experiencia ni interna, ni inmediata. Su manifestación fenoménica es tortuosamente indirecta. La única manera de llegar a *inferir* su organización es por un método semejante al que sugiere el comentario anterior de Cohen (1977): sentarse a la puerta de su reducto invisible, ver lo que entra y sale de él, y tratar de imaginar lo que sucede dentro, a base de hacer planos conceptuales de dicho reducto; planos cada vez más precisos y coherentes, y basados en una gama lo más amplia posible de entradas y salidas. De forma que, a diferencia de la vieja psicología introspectiva y a semejanza del conductismo, la psicología cognitiva es una psicología estímulo-respuesta (E-R), en cierto modo.

1.2.4. *La indeterminación de los modelos cognitivos*

Una peculiar psicología E-R en que lo que interesa es, precisamente el guión que hay entre la E y la R. Ese guión se compone, fundamentalmente de operaciones de cómputo sobre representaciones simbólicas (en el paradigma dominante, con lo que tenemos un nuevo guión C-R) o, en un sentido más laxo que sirve para todos los cognitivos, de procesos mentales que se aplican a estructuras de conocimiento. El modo de reconstruir tales procesos y estructuras es tan inferencial que se plantea una duda seria: ¿que certeza podemos tener los psicólogos cognitivos en relación con nuestros modelos?, ¿no serán esencialmente *indeterminados,* de tal forma que no sea posible, en último término, elegir entre modelos alternativos que predigan y expliquen igual de bien los datos? La sospecha de que quizá los modelos cognitivos puedan tener problemas importantes de indeterminación ha sido formalizada y defendida rigurosamente por John Anderson (1978). No es el momento de detenernos en esta demostración matemática, que ha sido cuestionada por otros investigadores, dando lugar a un interesante debate (Anderson, 1979; Pylyshyn, 1979; Hayes-Roth, 1979; vid. Rivière, 1986). Basta con abocetar lo que dice el argumento de Anderson: si los modelos cognitivos consisten en pares C-R (de cómputos sobre representaciones), siempre es posible que dos modelos diferentes predigan igualmente un cierto conjunto de fenómenos empíricos, con las condiciones siguientes: (1) que aquello que pueda representarse en el primer modelo pueda representarse —de diferente manera— en el segundo, (2) que los procesos sean realmente procesos de cómputo (que pueden entenderse como funciones recursivas). Independientemente del rigor matemático de la demostración de Anderson, expresa una inquietud frecuente —aunque no siempre confesada— de los psicólogos cognitivos sobre el valor de las inferencias que ponen en el «guión». Un guión que, en sus manos, pierde su simple y horizontal linealidad, para convertirse en una barroca estructura de procesos y representaciones.

1.2.5. *La conciencia del autómata abstracto: un nuevo fantasma en la máquina*

Ya hemos reparado (y volveremos a hacerlo) en el hecho de que no sabemos bien qué papel asignar a la conciencia en ese guión que dirige

el frío drama de las computaciones. Eso es fácil de comprender: el concepto de computación (que es más bien vago, vid. Johnson-Laird, 1988) sólo tiene sentido en el marco de la noción de *autómata* y, en ese paisaje conceptual, no hay sitio claro donde poner a la conciencia. James (1890) ya lo veía claro, en su tiempo, cuando se enfrentaba a las teorías mecanicistas: «¿Cual podría ser —decía— según este punto de vista la función de la propia conciencia? Si se trata de función *mecánica,* ninguna /../ sería una cadena de sucesos completamente autónoma, y cualquiera que fuera la mente que la acompañara no pasaría de ser un "epifenómeno", un espectador inerte, una especie de "aura, espuma o melodía", como dice Mr. Hodgson, que independientemente de que apoyara esos sucesos o se opusiera a ellos sería igualmente impotente para modificarlos» (p. 133), y, en otro momento: «si la conciencia es útil, tiene que serlo por su eficacia causal, y la teoría del autómata debe sucumbir al sentido común» (p. 147). La razón de las difíciles relaciones entre la mente fenoménica y la computacional es, como vemos, una razón de fondo: si hay una mente que «decide en conciencia», i.e. precisamente en virtud de ella, entonces esa mente no admite una descripción *completa* como autómata. Si la mente es, en el sentido ordinario de la palabra, un autómata, entonces la conciencia es un inútil y gravoso epifenómeno. Y esto va contra el sentido común. James se queda con el sentido común y con «lo que le dice su conciencia».

Pero después de cuarenta años de ciencia cognitiva esa posición «de sentido común» es difícil de aceptar. El *mecanicismo abstracto,* en que se basa la ciencia cognitiva, ha sido muy eficaz para comprender la mente: «los resultados alcanzados son extraordinarios —dice De Vega, 1984—; nunca se había dispuesto de tal cantidad de datos relevantes sobre los procesos mentales ni se había progresado tanto en su comprensión teórica» (p. 23). En los casos en que ciertos conceptos científicos, que son fértiles para la investigación y la comprensión teórica, no se someten al sentido común, los científicos convienen implícitamente en tomar una opción clara: peor para el sentido común. No puede decirse que la física cuántica, por ejemplo, sea una ciencia de sentido común. Al contrario: es claramente contra-intuitiva. Como dice Max Delbruk (1986), nuestra mente parece estar adaptada para afrontar la vida en un mundo de *dimensiones intermedias* (lo que Gerhard Vollmer, 1984, llama «mesocosmos») y sus intuiciones no resultan útiles cuando se dirige a lo mínimo e instantáneo (en el «microcosmos» de la física atómica y de partículas) o a lo inmenso y duradero (en el «macrocosmos» de la cosmología). De forma parecida, nuestra mente parece estar adaptada a un mundo en que resulta útil evolutivamente la intuición de la eficacia causal de la conciencia, pero ello no quiere decir que esa presuposición central del sentido común tenga que ser adoptada por la psicología cognitiva.

Las dificultades para el conocimiento objetivo de la mente son, por lo que llevamos dicho, claras:

1. La *mente fenoménica,* a la que tenemos acceso consciente, se enclaustra en una subjetividad de la que es muy difícil que salga. Así, los resultados de la introspección no llegan a adquirir fácilmente el estatuto *inter-subjetivo* que requiere la construcción de una ciencia objetiva. Wundt (1887) y James (1890) ya tenían una intuición clara de las dificultades, que finalmente llevaron al intento conductista de hacer de la psicología una ciencia que no hablara de la mente.

2. La recuperación del lenguaje mental ha venido, sobre todo, de la mano de los psicólogos cognitivos que *no* se refieren a la mente fenoménica sino a la *computacional* o, en sentido más amplio, a una mente que funcione, sin necesidad de «homúnculos» pseudo-explicativos, y que funcione precisamente en virtud de que posee estructuras de conocimiento o de representación.

3. Este plano de lo mental *no* se identifica descriptivamente con el anterior y, entonces, se plantea un complejo problema de relaciones. Si se acepta la metáfora del autómata, no resulta fácil ver cómo pueda ser compatible con la intuición de sentido común de la eficacia causal de la conciencia. Si no se acepta, se renuncia a una descripción mecanicista *completa.* Si la descripción mecanicista no es completa, deja de ser mecanicista.

4. Por estas y otras razones (vid. Jackendoff, 1987; Johnson-Laird, 1983, 1988), la conciencia no termina de encajar en el paradigma dominante en psicología cognitiva.

5. Finalmente, si la mente computacional *no* es inmediatamente fenoménica, tiene que ser reconstruida de forma muy indirecta, y las inferencias sobre lo mental plantean, en ese caso, un problema de indeterminación, independientemente de que éste se interpreterigurosamente como una limitación intrínseca del conocimiento sobre la mente (Anderson, 1979) o cómo una condición de todas las ciencias con un carácter inferencial acusado, como la física teórica (Pylyshyn, 1979).

1.2.6. *¿Demasiada mente para tan poca?*

La proposición primera, de la que partía nuestra reflexión sobre los fundamentos de la psicología cognitiva, era muy simple: es difícil desarrollar una ciencia qiie sea, al mismo tiempo, *mentalista y objetiva.* Algunas de las dificultades se han revelado en los párrafos anteriores; otras irán poniéndose de manifiesto a lo largo de nuestra indagación. Johnson-Laird (1983), de quien antes citábamos un comentario optimista sobre la posibilidad de una ciencia objetiva de la mente, también tiene sus momentos de duda:

> ¿Es posible la psicología? —dice— Quizá los psicólogos seamos los últimos en conocer la respuesta a esta cuestión. Raras veces nos la planteamos. Y, sin embargo, una conjetura plausible es que la mente tiene que ser más complicada que cualquier teoría que se proponga para explicarla: cuanto más compleja sea la teoría, más tendrá que serlo la mente que la pensó por primera vez. De ello se seguiría que la psicología está tratando de abarcar algo que le viene demasiado grande —está tratando de abarcarse su propia cabeza— y nunca podrá lograrlo. Esta idea es una variación del viejo chiste de que, en la democracia, los votantes siempre son más estúpidos que sus líderes. Cuanto más estúpidos son los líderes, más lo son los votantes por haberlos elegido. El chiste es obviamente falso; la observación psicológica más sutil e intrigante (op. cit., p. 1).

¿Son las dificultades anteriores expresiones de una dificultad intrínseca de la mente para conocerse *de forma completa* a sí misma? Pudiera ser que en psicología tuviéramos que contar con algo semejante —metafóricamente— al teorema de la incompletitud de Godel, que establece que ninguna lógica puede abarcarse a sí misma de forma completa y sirviéndose de su propio lenguaje (Arbib, 1964; Hofstadter, 1987). Acaso suceda que ninguna mente pueda abarcar su complejidad sirviéndose sólo de sus únicos recursos y empleando su propio lenguaje: el lenguaje de lo mental. Pero, ¿cómo reducir ese lenguaje intencional al puramente extensional que se emplea para aquellos objetos de los que no se predican atributos mentales?

Muchos filósofos de la mente se han planteado ese problema. Son también numerosos los que afirman, sencillamente, que es imposible realizar la traducción del lenguaje mental (compuesto de términos tales como «pensamientos», «creencias», «proposiciones», «imágenes», «representaciones», etc.) al lenguaje extensional que se emplea en las ciencias biológicas y físicas. Para algunos, esa imposibilidad no hace más que

demostrar que el lenguaje mentalista es un artefacto práctico para las interacciones entre humanos, pero no tiene cabida en ninguna clase de explicación científica, debiendo eliminarse en favor de un lenguaje escuetamente extensional, en que ya no existirían objetos como los pensamientos, los esquemas o las representaciones (Davidson, 1981; Quine, 1960; Rosenberg, 1980; Searle, 1983). Otros, por el contrario, piensan que la dificultad de traducción del lenguaje de lo mental al extensional es una demostración más de que el primero es *indispensable* para la ciencia cognitiva (Fodor, 1981).

En este momento del conocimiento científico de la mente es difícil decidir entre la alternativa de la eliminación del lenguaje mentalista y la de su conservación. Los desarrollos conexionistas sugieren la interesante posibilidad de concebir el significado de los términos mentales como refiriéndose a *propiedades globales,* emergentes del funcionamiento de sistemas de cómputo compuestos de unidades y asociaciones que funcionan de forma masiva y paralela, y que se describen de una forma puramente extensional (activaciones de las unidades, fuerzas de conexión entre ellas, umbrales necesarios para su out-put, etc). Según este punto de vista, los esquemas, los conceptos y las representaciones simbólicas del paradigma C-R no son más que formas metafóricas de designar esas propiedades que resultan del funcionamiento integrado y simultáneo de los sistemas de procesamiento distribuido y paralelo (Clark, 1989; Rumelhart, Smolensky, McClelland y Hinton, 1986). Pero ésta es una posibilidad de la que hablaremos más despacio en otro apartado.

Como vemos, hay razones fundadas para aceptar el humilde punto de partida que nos habíamos propuesto, a saber, que es difícil el intento de desarrollar una psicología del conocimiento que sea, al mismo tiempo, mentalista y objetiva. El psicólogo cognitivo se propone la difícil tarea de desarrollar una ciencia empírica de una mente que al mismo tiempo se auto-contiene y manifiesta, y se desconoce y oculta. Lo hace además con su propia mente que, por definición, no es más compleja que el sistema que trata de reconstruir. Y su reconstrucción se basa en inferencias muy indirectas, que dejan en el aire una duda sobre su posible indeterminación. Es un intento, el del psicólogo, claramente desmesurado.

Sin embargo, la desmesura del intento no es, en sí misma, ni un defecto especial ni una prerrogativa exclusiva de la psicología cognitiva. Todas las ciencias tienen un objetivo final desmesurado: conocer la realidad. Los psicólogos arduamente formados en el operacionalismo, un

poco estrecho, de los años sesenta, sentimos un cierto escalofrío cuando un físico como Hawkins (1988) nos dice: «El objetivo final de la ciencia es proporcionar una única teoría que describa correctamente todo el universo» (p. 29). Es éste, sin duda, un objetivo sin mesura. Pero las ciencias son mesuradas en sus objetivos inmediatos y procedimientos: acotan y recortan los problemas hasta hacerlos tratables, miden lo que pueden medir, limitan sus ambiciones provisionales a los problemas y dejan de lado los misterios (como parece ser el de la conciencia), es decir, aquellas cuestiones que, aun siendo intrigantes, no son conceptualmente claras o accesibles a una indagación empírica a corto plazo. Estas son también las cautelas que empleamos los psicólogos cognitivos para dirigirnos, mesuradamente, a nuestro objetivo final desmesurado: conocer la mente.

Sección segunda:
LOS OBJETOS CON MENTE

Capítulo 2

LOS OBJETOS CON MENTE Y LA METAFORA DEL ORDENADOR

2.1. La categoría natural de los objetos con mente

Con independencia de los obstáculos para acceder a un conocimiento objetivo de la mente, lo cierto es que las personas hacemos una división natural del mundo en dos amplias categorías de objetos: de unos pensamos y decimos que tienen mente; a otros no se la atribuimos. Intuitivamente, predicamos lo mental cuando decimos que algo *piensa, percibe, desea, recuerda, siente.* Siempre que empleamos esos verbos para referirnos a algo, estamos atribuyendo mente. No es inexacto decir que lo que hacemos los psicólogos, en esencia,, consiste en trabajar con esos verbos: con sus significados y referentes. Tratamos de conocer de forma científica en qué consisten esas funciones de pensar, desear, etc, que se predican de ciertos objetos y no de otros. Nos interesan los objetos de los que se dicen funciones mentales (Fm), como las personas y los animales, y no dedicamos nuestra atención a otros muchos de los que no se dicen (-Fm), como las plumas estilográficas y las estrellas del cielo. Nuestro campo de investigación puede definirse, así, de manera muy escueta: Fm; el estudio de -la naturaleza, estructura, génesis y funcionamiento de las funciones mentales. Esto se aplica también a aquellos psicólogos que, finalmente, llegan a la conclusión de que no puede hacerse ciencia con un lenguaje mentalista, pues a ellos también les interesa explicar en qué consisten funciones tales como pensar, recordar y tener emociones.

Por todo ello, no deja de ser sorprendente la escasa atención que tradicionalmente hemos dedicado los psicólogos a la semántica y el uso cotidiano de los verbos Fm, y nuestra tendencia a considerar como «cosa

de filósofos» o, en todo caso, «de lingüistas» la indagación de sus características peculiares, estructura y uso en el lenguaje, así como de las propiedades lógicas especiales que tienen los enunciados con verbos mentales. En filosofía hay una poderosa y ya vieja tradición de investigación y debate sobre el lenguaje psicológico (Carnap, 1954; Chisholm, 1957; Church, 1954; 1956; Davidson, 1965; Putnam, 1981; Quine, 1960; Simpson, 1973, 1975; McCauley, 1987; Fodor, 1988; Searle, 1983) que los psicólogos ignoramos con la misma tenacidad con que se ha desarrollado.

Este residuo localista de nuestra larga y difícil separación de la filosofía ya no es justificable. El conocimiento de la larga tradición de pensamiento sobre lo mental desarrollada en diversas perspectivas filosóficas puede ayudarnos a los psicólogos en dos aspectos importantes: en primer lugar, puede contribuir a una mejor comprensión de nuestro objeto de estudio que, con frecuencia, se considera de forma un poco ingenua y acrítica; además, puede facilitar el estudio de problemas propiamente psicológicos, pero que han sido descuidados tradicionalmente, tales como los relacionados con las representaciones cognitivas de lo mental y las habilidades de «cognición psicológica» que las personas usamos en las interacciones cotidianas. En las concepciones teóricas que los psicólogos hemos desarrollado acerca del pensamiento y la inteligencia, por ejemplo, la huella dejada por la tendencia a presuponer que esas funciones se ejercen, esencialmente, «en frío» y sobre objetos «apersonales», resulta muy visible en el cuerpo conceptual de la psicología.

El propósito de esta sección, y la siguiente, es el de profundizar en nuestra comprensión de los fundamentos de la psicología cognitiva, a partir de un análisis de los usos lingüísticos, las «categorías naturales» (Rosch, 1978; Rosch y Mervis, 1975; Mervis y Rosch, 1981) y los procesos de razonamiento e inferencia sobre lo mental. En esta sección nos fijaremos sobre todo en los objetos de los que se predican funciones mentales, para realizar desde ahí una reflexión sobre la metáfora del ordenador. En la siguiente, analizaremos las características de tales funciones. En ambas secciones, se entreverán, con la liberalidad que el rigor permita, observaciones históricas, razones lógicas, consideraciones filosóficas, intuiciones cotidianas y concepciones psicológicas.

Empecemos por una constatación muy simple: la categoría constituida por los objetos a los que se aplican verbos mentales no es como las clases lógicas bien formadas, en que todos los elementos tienen el mismo grado de pertenencia, y los límites son claros e inequívocos. Tiene, por el contrario, límites difusos y una estructura distribucional semejante a

la de otras categorías naturales, cuyos elementos no son estrictamente equivalentes, sino desigualmente representativos. Del mismo modo que los gorriones son más representativos —más prototípicos— de la categoría natural de las aves que los pingüinos (Malt y Smith, 1984), las personas lo son más de la categoría Fm (de los objetos a los que se atribuyen funciones mentales) que las bacterias. De la personas decimos que *perciben, piensan, recuerdan, desean,* etc, pero no está claro cuáles de estas funciones —si es que alguna— podrían aplicarse a las bacterias.

En realidad, las personas son *prototipos* de la categoría de objetos (Fm). Lo son menos los animales, y tanto menos cuanto menos complejos son y más se alejan del *filum* humano. Algunos de los rasgos, que definen por comprensión la categoría natural de lo mental (por ejemplo, la percepción), se aplican frecuentemente a todos ellos; otros, como el pensamiento, a muy pocos o solo al hombre (me refiero al uso cotidiano en el lenguaje de la categoría natural de lo mental, y no al tratamiento científico del problema del pensamiento animal, tal como se plantea, por ejemplo, en Angüera y Vea, 1984, Premack y Premack, 1983 o Griffin, 1986). En el polo opuesto, el de la categoría (-Fm), tenemos objetos tales como las piedras, los planetas y sus satélites, los tenedores y (con un grado menor de representatividad) las plantas.

En nuestra cultura, y entre adultos, los rasgos Fm no suelen áplicarse a este tipo de objetos a no ser metafóricamente y en usos poéticos, como cuando Aleixandre dice: «Sólo la luna sospecha la verdad. /Y es que el hombre no existe»; o García Lorca: «La viuda de la luna/ ¿quién lo olvidará?/ Soñaba que la tierra/ fuese de cristal». Es obvio que ni Aleixandre ni García Lorca *creen,* literalmente, que la luna *sospeche* o *sueñe.* Estos son, de nuevo, rasgos de lo mental y que definen intensionalmente la categoría Fm. Precisamente, la fuerza expresiva de las metáforas de Lorca y Aleixandre se deriva de la atribución animista de verbos (Fm), que suscitan emociones con mayor poder y facilidad que los (-Fm).

En los usos cotidianos de nuestra cultura, las funciones Fm se predican, en especial, de objetos que poseen dos propiedades: ser sistemas vivos y complejos. En último término, las propias personas poseen estas propiedades, y la atribución de funciones mentales, en nuestra especie, parece cumplir fundamentalmente la función adaptativa de predecir y explicar la conducta de los congéneres. Son éstos, como hemos dicho, los prototipos indiscutibles de la categoría de objetos Fm (i.e. de objetos de los que se predica que piensan, comprenden, recuerdan, perciben, desean, etc). A estos prototipos se les aplican todas las funciones Fm.

Si definimos por comprensión la categoría **Fm,** por los *rasgos* que contiene (tal como se hace en muchos modelos cognitivos sobre los conceptos; vid., por ejemplo, Sainz, 1985, y Smith, 1989), podemos decir que tales rasgos son sencillamente el conjunto de las funciones que se entienden naturalmente como «mentales»: percibir, creer, pensar, desear, saber, recordar, intuir, etc. Los verbos mentales son, así, los rasgos que definen intensionalmente la categoría (su imprecisa definición extensional incluye habitualmente objetos tales como las personas, los chimpancés y las jirafas). Pues bien, el universo total de los rasgos (tanto definitorios como característicos, si cabe tal distinción; vid. Smith, Soben y Ripps, 1974; Smith y Medin, 1981; Rosch y Mervis, 1975), que definen la categoría Fm, se predica de su elemento prototípico por excelencia: las personas. Las funciones mentales que atribuimos a las personas agotan intensionalmente la categoría de lo mental.

En realidad, un supuesto básico subyacente a todas las interacciones entre humanos es la noción de que los otros son seres con una mente cuya estructura esencial es idéntica a la propia. Sin embargo, dejando de lado esos prototipos centrales de la categoría Fm, los objetos que se sitúan más en sus límites resultan más dudosos y, sobre todo, resulta dudoso *qué* funciones Fm pueden atribuirse a ellos (un psicólogo, al que este trabajo debe mucho, me refería recientemente la respuesta de su hija a la pregunta de si los perros piensan: «Piensan, sí. Pero lo que no tienen son ideas»).

Una ilustración muy significativa del carácter difuso de la categoría de los objetos a los que se atribuyen funciones mentales es el hecho de que, en la historia del pensamiento, se ha argumentado la necesidad *literal,* (y no ya la conveniencia metafórica) de atribuir funciones Fm a toda la realidad natural. Recordemos, por ejemplo, que Leibniz, en la *Monadología* (1720; vid. ed. de 1981) sostenía la idea de que las mónadas, «los verdaderos átomos de la naturaleza y, en una palabra, los elementos de las cosas» (p. 73) poseían un principio interno, una actividad intrínseca, de *percepción.* En esta interpretación la función Fm «percepción» es constitutiva, en último término, de toda la realidad. Una idea no muy lejana de la que sostenía Fechner (1846), que argüía —en contra del materialismo mecanicista de la ciencia de su tiempo— la necesidad de afirmar que las plantas y, finalmente, todos los objetos naturales, poseen una vida mental (Fechner, 1851). Como ha señalado Boring (ed. 1988), esa posición no contribuyó precisamente a facilitar la aceptación de la obra de Fechner en los círculos científicos.

2.2. El animismo del pensamiento natural y el mecanicismo de la ciencia natural

Por otra parte, la *extensión* de la categoría Fm varía: (1) a lo largo del desarrollo de las personas, y (2) de unas culturas a otras. El concepto de «animismo» hace referencia a la tendencia que tienen los niños y miembros de otras culturas a atribuir funciones Fm (y vida) a objetos que no se incluyen en la categoría Fm de los adultos occidentales. Es lo que sucede, por ejemplo, en el caso de «Kent», uno de los niños interrogados por Piaget en su estudio sobre *La representación del mundo en el niño:* «Si pinchamos esta piedra, ¿lo sentirá? —No. —¿Por qué? —Porque es dura. —¿Lo sentirá si la ponemos en el fuego? —Sí. —¿Por qué? —Porque esto la quema» (1926, cit. ed. esp. de 1984, p. 157). En este caso, se atribuye la función mental «sentir» a uno de los objetos que son, para el adulto occidental, prototípicos de la categoría (-Fm): una piedra. El animismo en el niño no es un simple producto lúdico o el resultado de una actitud metafórica, sino una expresión profunda de una propensión epistémica a confundir objeto conocido y sujeto que conoce (DelVal, 1975) y quizá —como iremos viendo a lo largo de estas páginas— una expresión aún más profunda de algo más: un «efecto distorsionante», sobre el conocimiento de lo real, del funcionamiento de un sistema cognitivo especialmente dotado para comprender y atribuir estados mentales (Astington, Harris y Olson, 1988; Cosmides, 1989).

Con independencia de las críticas que puedan hacerse a las *interpretaciones* de los fenómenos animistas (vid., por ejemplo, Geertz, 1973), lo que no se cuestiona es su existencia y consistencia como tales fenómenos: son muy universales en las culturas y se enraizan con fuerza en el pensamiento, tanto del niño como del hombre que no ha interiorizado a fondo la revolución conceptual de la ciencia que nació en el Renacimiento. De hecho, uno de los obstáculos epistemológicos más impenetrables y persistentes que ha tenido que superar el conocimiento, para la constitución de una ciencia objetiva de la naturaleza, ha sido la tendencia a interpretar como intencionales fenómenos que admiten una descripción completa en términos extensionales.

En los términos analíticos que estamos empleando, un efecto importante de la revolución científica sintetizada en el nombre de Newton ha sido éste: la extensión de la categoría natural (Fm) se reduce considerablemente en aquellos individuos que realizan, en su propia cabeza, una profunda reconstrucción conceptual. Una transformación semejante a la

que permitió la creación de las ciencias de la naturaleza, mediante operaciones sucesivas de limpieza de los residuos animistas de su campo conceptual. Como dice Angel Pérez Gómez (1985), cada niño en su escuela tiene que realizar una ruptura epistemológica (más bien varias) para acceder a las formas del pensamiento científico; debe someter sus conceptos implícitos a una tarea de reelaboración y sustitución (Pozo, 1987). Una parte esencial de esa tarea es, precisamente, la que implica eliminar adherencias animistas y reducir mucho el universo de objetos de los que se predican funciones Fm.

La superación de las propensiones animistas ha sido el producto de un largo esfuerzo del pensamiento humano: un esfuerzo que ya era claro en los primeros filósofos naturalistas griegos y que alcanza un punto de inflexión con la creación de la «nuova scienza» del Renacimiento y la obra de Newton. Mientras que, en el pensamiento cosmológico de los griegos y de la Edad Media, las adherencias animistas seguían estando presentes (recordemos, por ejemplo, cuáles eran los principios de todas las cosas para Empédocles: tierra, agua, aire, fuego..amor y odio), la gran revolución conceptual de Newton gravitó sobre un punto esencial: las leyes de la cosmología no añaden nada cualitativamente distinto a las de la mecánica, en la que no cabrían categorías (Fm). Como señala Aracil (1986):

> De la confluencia metodológica entre el estudio del movimiento de los astros en el firmamento, y el de las máquinas, aquí en la tierra, surge la ciencia de la mecánica. Con ella se desencadena el primer programa coherente de interpretación del mundo físico. Empieza así su andadura una de las realizaciones más características del hombre moderno: la ciencia física. Se inicia, por citar un nombre, con Galileo, y alcanza sus mayores cotas de brillo y esplendor con Newton, Lagrange y Laplace. Su nota característica fundamental es el *deterninismo* (p. 48).

La genialidad de Newton consistió, en esencia, en descubrir la identidad de fondo, que subyacía a dos dominios previamente separados de conocimiento: el de las trayectorias de los cuerpos celestes y el de los fenómenos mecánicos en relación con proyectiles, bolas, etc, que antes de él había estudiado Galileo. Así, por ejemplo, comprendió que la curvatura de la trayectoria de la luna —uno de los objetos (-Fm), de los que hemos hablado antes— seguía la misma pauta matemática que la de los objetos terrestres, (-Fm), cuyas trayectorias había analizado Galileo, y sintetizó esa pauta común en una Ley de Gravitación Universal. Toda

esa gigantesca operación conceptual tuvo su base en una consideración mecanicista y estrictamente extensional de los fenómenos celestes, para cuya interpretación aún se empleaban categorías teñidas de intencionalidad en el Renacimiento. La interpretación rigurosamente mecanicista del cosmos permitió formular leyes universales y de gran elegancia matemática sobre el comportamiento de los cuerpos celestes. Como dice también Aracil (op. cit.),

> el deslumbramiento que produjeron estos resultados en las mejores mentes de la época es incalculable. Se comprendió que, al menos en un determinado dominio de nuestra experiencia, era posible resumir la enormidad de datos de la experiencia mediante leyes de gran sencillez y de formulación matemática. Inmediatamente se percibió que uno de los proyectos más sugestivos a que podía dedicarse el hombre de pensamiento era el buscar otros dominios de la experiencia a que fuese aplicable el mismo método (p. 46).

La eliminación de los insidiosos residuos animistas para la comprensión del mundo físico —que, en realidad, no se logra hasta Galileo y Newton— establecía una escisión tajante entre los enunciados extensionales de la ciencia, susceptibles de tratamiento matemático y basados en la observación de fenómenos, y los de carácter intencional que se realizan cada vez que se dice que un algo *piensa, comprende, percibe, recuerda,* etc. Es importante tener presente el hecho de que esa escisión se cimentó en una consideración *mecánica* de la naturaleza. Desde ella, se planteaba ese desafío fascinante de ir ganando para la consideración extensional y mecanicista dominios cada vez más amplios de conocimiento.

Tales son las coordenadas que permiten comprender los dos grandes programas de investigación que darían origen a la psicología moderna. El de los empiristas estaba atrapado en una reducción inevitable, si querían ser consecuentes con sus propios presupuestos: la reducción de las Fm a la *mente fenoménica,* la reducción —en otras palabras— de la mente a conciencia. Desde ese axioma reductivo, se planteaba una pregunta esencial que podemos traducir así: ¿es posible descubrir, en el mundo privado de las ideas de la mente, de las Fm, pautas de orden semejantes a las descubiertas a partir de la consideración puramente extensional, mecanicista, de lo natural?; ¿pueden formularse las «leyes de gravitación universal de la mente»? El mecanismo de asociación entre ideas —concebidas como fenómenos inmediatamente observables en la mente— parecía proporcionar el mecanismo buscado.

Por su parte, el programa racionalista, bajo la sombra del muro infinito levantado por Descartes para separar las realidades extensionales de las intencionales, sólo tenía un camino para realizar el desafío de conquistar dominios nuevos para la visión mecanicista y extensional: la concepción de los animales como mecanismo complej simos, como seres sin conocimiento ni mente, como «máquinas hechas por las manos de Dios», en palabras de Descartes (1637, p. 72 de la ed. esp. cit. de 1989).

2.3. La categorización de lo mental en la psicología cotidiana

En todo caso, la herencia dejada por dos mil quinientos años de lucha por comprender la naturaleza, y en especial el legado de la revolución paradigmática de Newton, dejaban al hombre moderno en una situación relativamente confortable en relación al problema de la delimitación de los objetos a los que podrían aplicarse funciones Fm: serían, en último término, objetos *vivos,* organismos. Fueran lo que fueran las funciones mentales, eran funciones de un organismo. Lo que ya no estaba tan claro era su naturaleza como tales funciones. La mente no encajaba del todo en la concepción radicalmente extensional (originada en la consideración mecánica de la naturaleza) que resultó de la importante herencia de Newton, y, como señalaba Johnson-Laird (1988) en un comentario ya citado: «se consentía que los adultos pudiesen hablar de ella en privado, siempre y cuando comprendieran que, en realidad, no existía» (p. 13). Cuando menos, se tenía la garantía de que la categoría (-Fm), formada por máquinas y otros seres sin vida, no incluía esa elusiva entidad, difícilmente objetivable, de la que podía hablarse en privado, pero no en voz alta en los ámbitos públicos de la ciencia: la categoría (-Fm) podía explicarse en términos exclusivamente extensionales.

Desde estas consideraciones puede entenderse bien la importancia *psicológica* del desarrollo de la tecnología del conocimiento: antes contábamos con una cómoda distribución de papeles que permitía salvaguarı dar de la dudosa y privada categoría Fm, cuando menos, a los artefactos creados por el hombre y a la naturaleza inorgánica; restringir la peligrosa influencia de los predicados Fm. Tenedores, planetas, nubes y relojes se salvaban de la inquietante asignación a una categoría tan eminente en la superficie como incomprensible en el fondo.

En cuanto a las funciones Fm de las personas y otros organismos complejos, cabían esencialmente tres posibilidades:

(1) Realizar una reducción extensional completa de ellas (la posición «objetivista» de Watson), desde una psicología «en tercera persona de singular» y de base mecanicista.

(2) Lidiar penosamente con la desalentadora pretensión de objetivar los datos de conciencia (psicología introspectiva), reduciendo quizá las posibilidades de la introspección al estudio de las funciones Fm más cercanas al mundo extensional, como la percepción (ésta era, en esencia, la posición de Wundt; vid. Blumenthal, 1975; Danziger, 1980; Leary,1979; Miraíles, 1986; Pinillos, 1981), y dejando las otras al cuidado de la psicología social.

(3) Desarrollar una ciencia más liberal de la mente, no basada esencialmente en la introspección, y definiendo sus propias leyes y formas de organización como *no-reductibles.* Esta posición se basaba en el supuesto de que, en el ámbito de lo mental, emergen propiedades conñguracionales nuevas, predicables de objetos intencionales (las formas perceptivas, los esquemas, las operaciones, etc). Estas constituirían bien una prolongación de funciones adaptativas del organismo (Piaget) o espejos paralelos —pero no tangentes— de leyes de campo del sistema nervioso *(gesfalten).* Si bien este enfoque ya estaba cerca de lo que luego ha sido el proyecto de una psicología cognitiva, siempre dejaba en el aire la duda de cómo justificar, *de forma tangible* y sobre supuestos mecanicistas —en el mejor sentido de la palabra—, la emergencia de funciones intencionales (la mente) de objetos extensionales (los organismos).

2.4. El desafío de Turing

Tal es, descrita de forma sucinta, la situación en que hace su aparición un nuevo elemento que vuelve a tambalear e indiferenciar los límites imprecisos de la categoría Fm: la tecnología del conocimiento. Antes de ella, por lo menos estaba claro algo: que, con perdón de la redundancia, hay *cosas* y *no-cosas.* Cosas tales como las montañas, las sillas, los cuchillos de cocina, las nubes y los automóviles que lo son en sentido estricto, ya que la categoría natural de «las cosas» venía a identificarse, *grosso modo,* con esa que nosotros hemos llamado de forma más concisa: (-Fm), aquello de lo que no se predica mente. Por otra parte, estaban los objetos «no-cosas», (Fm): los organismos capaces de realizar actividades de percepción, recuerdo y —en su caso— *pensamiento,* la función más eminente y prototípica de las no-cosas; una función que, en las explicaciones más coherentes y profundas (Dewey, 1933; Piaget, 1961)

se entendía como una prolongación de funciones biológicas de adaptación y organización que poseen los organismos precisamente en tanto que tales.

La tecnología del conocimiento implica, en primer lugar, una ruptura del *statu quo* alcanzado con relación a las categorías Fm y -Fm. En segundo lugar, proporciona un nuevo impulso al intento de conocer la mente en términos precisos, mecanicistas, respetuosos con sus propiedades «emergentes», y estrictamente objetivos. Ello se debe, sobre todo, a que, como ha señalado recientemente Varela (1988), la tecnología cognitiva proporciona al hombre una nueva imagen de sí mismo, de su mente: La imagen de una maquinaria que maneja representaciones simbólicas o subsimbólicas —*computa* representaciones— y, de forma automática, produce efectos que se parecen mucho, en aspectos relevantes, al pensamiento humano. Si se admite esta metáfora, podemos decir que el hombre tiene, en los últimos años, la sensación apasionante de que está produciendo funciones mentales (Fm) con sus técnicas. O, como mínimo, de que está produciendo «cogito», pensamiento, que ha sido la marca de fábrica de lo mental por lo menos desde el origen de la epistemología moderna con Descartes.

La tesis de la máquina que produce pensamientos fue defendida, de forma brillante y provocativa, por uno de los más inteligentes precursores de las ciencias cognitivas: Alan Turing. El título de su artículo clásico sobre el tema era «Computing Machinery and Intelligence» (1950). Turing planteaba el asunto del siguiente modo: supongamos que, en vez de dedicarnos a especulaciones subjetivistas sobre qué es pensamiento y qué no, hacemos un juego que incluye una *prueba objetiva.* El juego consiste en lo siguiente: se crea una situación en que el que administra el test hace preguntas a seres que están en otra habitación y a los que, por consiguiente, no puede ver ni tocar. El método para hacer las preguntas puede consistir, por ejemplo, en dárselas escritas a un mensajero que se encarga de llevarlas hasta la habitación donde están los seres interrogados y de traer de vuelta las respuestas dadas por esos seres, escritas en papelitos. Supongamos también que esos seres son dos: un hombre, B, y una máquina, A. Si un observador imparcial fuera incapaz de distinguir —sin verlos ni tocarlos— al hombre B, al que interroga sobre sus pensamientos, de la máquina, A, a la que pregunta sobre los suyos, ¿con qué derecho podríamos decir que el hombre piensa y la máquina no? Si la máquina *imita* perfectamente al hombre en sus manifestaciones cognitivas, ¿por qué no decir también que «piensa»?:

> Hemos sugerido sustituir la pregunta «¿pueden pensar las máquinas?» por la de «¿existen computadoras digitales imaginables que jueguen bien el juego de la imitación?» Si se desea, puede generalizarse esa pregunta: «¿Hay máquinas de estado discreto que hagan un buen juego?» Pero, dada la propiedad universal (de las máquinas) vemos que ambas preguntas equivalen a: «Supongamos una determinada computadora digital C. ¿Es cierto que, modificando esta computadora para que tenga un almacenamiento adecuado y dotándola de un programa apropiado, podemos conseguir que C desempeñe eficazmente el papel de A en el juego de la imitación y el papel de B lo haga un hombre? (Turing, 1950, ed. esp. cit. de 1984, p. 24).

La respuesta de Turing a la pregunta de si las máquinas son *potencialmente* capaces de pasar con éxito su test sobre la función prototípica Fm de pensar es clara: sí, lo son *por principio.* Lo son en tanto que poseen una propiedad que las define esencialmente: la de ser universales, i.e. capaces en potencia de computar cualquier algoritmo bien definido, cualquier *procedimiento efectivo.* El hecho de que a las máquinas actuales se les resistan algunos cómputos no es relevante, en el marco de la reflexión de Turing. Porque él no se refiere a los ordenadores materiales, construidos hasta ahora —con todas sus limitaciones de memoria, software, etc— sino a una máquina abstracta de carácter universal, de la que las máquinas concretas no son más que realizaciones parciales.

Es importante que profundicemos tanto en la importancia del *desafío* que plantea Turing (1950), como en la significación real de la *respuesta* a ese desafío. En cuanto a lo primero, el juego al que, en realidad, quiere jugar Turing (y del que el otro no es más que una realización parcial, como lo son las máquinas particulares de su máquina universal) consiste en enfrentarse a la tarea conceptual de desdibujar de nuevo los contornos precisos de la distinción, tan arduamente lograda por la ciencia, entre las categorías (Fm) y (-Fm). Puede entenderse, si se quiere, ese juego como una metáfora provocativa (y el estilo del artículo de Turing nos muestra hasta qué punto se divertía con su travesura) pero, al mismo tiempo, como una metáfora profunda y capaz de movilizar el pensamiento: basta con invertir su dirección para reconocer su verdadero sentido. Este no era tanto, «¿pueden pensar las máquinas?», como «¿si las personas *también* piensan, qué hay de misterioso en esa actividad?» En una palabra: la formulación de Turing llevaba a plantear la posibilidad de una consideración estrictamente mecanicista, despojada de todo misterio, del pensamiento y era la primera expresión de lo que se ha llamado, en psicología cognitiva, «la metáfora del ordenador».

Pero conviene que nos detengamos algo más en la respuesta de Turing (1950): por principio, siempre que los «pensamientos» de las personas puedan entenderse como resultantes de *procedimientos efectivos* (i.e. de algoritmos bien definidos) podrán ser imitados por una máquina que tiene, precisamente, las características de ser (1) muy simple y (2) universal. Turing había imaginado y justificado tal máquina abstracta en un artículo anterior (1936), que se considera un clásico de la teoría de la computación. La máquina consistiría sencillamente en un dispositivo de estados finitos, equipado con una especie de cinta móvil. La cinta estaría dividida en casillas, cada una de las cuales podría contener un dígito con arreglo a un código binario. El movimiento de la cinta estaría regulado por un dispositivo capaz de «leer» los dígitos de las casillas y borrarlos o modificarlos. Esa máquina secuencial tendría un número limitado de estados, y éstos serían de carácter discreto, definiéndose tales estados por las «posiciones» anteriores y posteriores a cada una de las operaciones. Las reglas de actuación de la máquina serían muy simples: cambiar los 1 a 0 en las casillas, y viceversa, mover la cinta una casilla a la izquierda o una a la derecha y...pararse. La aplicación de tales reglas estaría controlada por un programa de instrucciones en la forma condición-acción. Las condiciones vendrían definidas por los símbolos de las cuadrículas de la cinta y los estados de la máquina; las acciones serían las cinco operaciones simples de pasar 1 a 0, 0 a 1, ir a la izquierda, ir a la derecha, y parar.

Turing (1936) inventó este otro «juego» en 1936 para probar algunos resultados, bastante abstrusos de la lógica simbólica. Demostró que todo aquello que puede ser descrito en términos de procedimientos efectivos puede también ser computado por una máquina tan simple como la descrita en el párrafo anterior. La única condición es «no escatimar» nunca cinta a la máquina, puesto que la cinta es, al mismo tiempo, la memoria, el input y el output con que cuenta. Si hay algo que, siendo un procedimiento efectivo, no pueda ser computado por la máquina, será porque no se le ha permitido emplear la cinta (la memoria) que necesitado bien porque no se han dado bien las instrucciones (programa). En las máquinas de Turing, tanto las instrucciones como los «datos» se simbolizan con cifras binarias, empleando un solo código-tipo para los estados de la máquina y las operaciones posibles sobre la cinta.

Así, cada máquina específica de Turing puede simbolizarse, simplemente, por una sola cifra binaria. La máquina universal puede simular las operaciones de cualquier máquina de Turing particular: lo que hace es leer la cifra que define a una máquina concreta y aplicarla a unos datos. Así, la máquina abstracta y universal de Turing es antecesora muy directa de los ordenadores. El marco conceptual desarrollado por Turing (separación de instrucciones y datos, empleo de un código binario, definición de una máquina de estados discretos, etc) fue el que recogió después John von Neumann, para aplicar el esquema lógico de Turing a materiales físicos concretos: de ahí nacieron los ordenadores (vid. Bolter, 1984). Un ordenador puede entenderse como una realización física de la máquina universal de Turing.

2.5. La simplicidad de la razón universal

La máquina de Turing es uno de los objetos menos misteriosos concebidos por la mente del hombre. Su expresión visible es elemental: una cadena de unos y ceros. Lo que esa expresión significa también lo es: algo tan simple y aparentemente poco inteligente como pueda serlo un autómata literalmente «a-mental», que se limita a cumplir, de forma inflexible y acrítica, sus instrucciones, sin poner nada de su parte. ¿Por qué tenía entonces tanta importancia la demostración de Turing, hasta el punto de volver a borrar los contornos de las categorías sobre lo mental, cristalizadas desde la genial reducción mecanicista del mundo extensional por Newton?

El secreto está en la propia simplicidad y en la ilimitada generalidad que, al mismo tiempo, tiene su poder de cómputo: en la fascinante propiedad que posee la máquina de ser precisamente un *autómata,* pero un autómata *universal,* capaz de aplicar cualquier clase de algoritmo. Quizá no resulte tan obvia, a primera vista, la importancia que tiene esa característica de universalidad, de modo que recurriré a una pequeña ilustración histórica que pone claramente de manifiesto esa importancia: cuando Descartes (1637), en *El discurso del Método,* trata de demostrar que los animales son máquinas, autómatas sin razón ni pensamiento, recurre a dos argumentos principales para esa demostración:

(1) «El primero de los cuales —dice Descartes— es que jamás podrán utilizar palabras, ni otros signos para comprenderlas, como nosotros hacemos para declarar a los demás nuestros pensamientos.»

(2) «Y el segundo es que aunque hagan muchas cosas igual de bien, o quizá mejor que alguno de nosotros, carecerían infaliblemente de otras, *por lo que se descubriría que no obran por conocimiento, sino por la*

disposición de sus órganos. Porque mientras que *la razón es un instrumento universal* que puede servir en toda clase de situaciones, esos órganos necesitan una disposición particular para cada acción particular, de donde deriva que es moralmente imposible que haya los suficientes en una máquina para hacerla obrar en todas las situaciones de la vida, de la misma forma que nuestra razón nos hace obrar» (pp. 73-74 de ed. esp. de 1989, el subrayado es nuestro).

Los animales, por tanto, *no obran por conocimiento,* no pertenecen, en realidad, a la categoría (Fm) que hemos descrito antes. La reducción mecanicista del mundo extenso les afecta por entero, lo cual, dice Descartes, «no parecerá nada extraño a quienes sabiendo cuántos autómatas diversos, o máquinas en movimiento puede hacer la industria de los hombres sin emplear más que unas pocas piezas /.../ consideren este cuerpo como una máquina que, hecha por las manos de Dios, está incomparablemente mejor ordenada y tiene en sí movimientos más admirables que ninguna de las que puedan ser inventadas por los hombres» (op. cit., p. 72). Lo que salva al hombre de esa reducción mecanicista completa es precisamente la presencia de una *sustancia* enteramente irreductible, el pensamiento, el origen último de todo aquello de lo que se puede predicar que es intencional. En último término, la universalidad de la razón es la marca final de esa sustancia intencional.

Dado que el pensamiento —esa sustancia mental— no es extenso ni tiene partes, no puede dividirse. Es imposible que, en esas condiciones, pueda ser considerado en términos mecánicos. Sólo de los animales que «obran por conocimiento» (los hombres) puede predicarse mente. Por su parte, Leibniz, a pesar de ampliar el concepto de mente (entendido en términos del rasgo conceptual (Fm) «percepción») a toda la naturaleza, sigue entendiendo que lo mental es irreductible a cualquier clase de explicación mecánica:

> Hay que reconocer, por otra parte, —dice en la *Monadología*— que la *percepción* y lo que de ella depende resultan *inexplicables por razones mecánicas,* esto es, por medio de las figuras y los movimientos. Porque, imaginémonos que haya una máquina cuya estructura la haga pensar, sentir y tener percepción; se la podrá concebir agrandada, conservando las mismas proporciones, de tal manera que podamos entrar en ella como en un molino. Esto supuesto, una vez dentro, no hallaremos sino piezas que se impelen unas a otras, pero nunca nada con lo que explicar una percepción. Así, pues, esto hay que buscarlo en la sustancia simple, no en lo compuesto de la máquina. Más aún, no cabe hallar en la sustancia simple otra cosa excepto esto, es decir, excepto las percepciones y sus cambios. Y solamente en esto pueden consistir todas las *Acciones internas* en las sustancias simples (ed. cit., 1981, p. 83).

La ontología de Leibniz es, ciertamente, muy diferente de la de Descartes: mientras que éste redujo a su mínima expresión lo que hemos llamado «categoría (Fm)» para dejarla resumida en un solo elemento natural —el hombre—, Leibniz, por el contrario, la amplía indefinidamente hasta hacerla abarcar a toda la naturaleza: en la sustancia simple, de la que todo se compone, no cabe hallar sino las percepciones y sus cambios. Esa diferencia se relaciona muy estrechamente con el tipo de funciones Fm que eligen uno y otro como prototipos y puntos de partida de su reflexión: el pensamiento (Descartes) y la percepción (Leibniz); también con una operación esencial que realiza Leibniz y que resultaba muy extraña al pensamiento cartesiano. Mientras que éste había tomado como lugar de origen la evidencia fenoménica consciente del cogito, y la identificación de las funciones Fm con la conciencia, Leibniz realiza una ruptura de enorme trascendencia (por ejemplo, para la propia psicología cognitiva), consistente en desvincular mente y conciencia:

> El estado transitorio que envuelve y representa una multitud en la unidad o en la sustancia simple —dice— no es sino eso que llamamos *Percepción,* que debemos distinguir de la apercepción o de la conciencia, como luego quedará de manifiesto. Y precisamente en este punto los cartesianos han caído en un grave error, por no haber tenido para nada en cuenta las percepciones de las que no nos apercibimos. A causa de esto también han creído que solamente los Espíritus eran Mónadas y que no había Almas en las Bestias ni otras Entelequias; y han confundido, junto con el vulgo, un largo aturdimiento con una muerte en sentido riguroso (ibid, p. 83).

Pero, a pesar de estas importantes diferencias ontológicas, Descartes y Leibniz coinciden explícitamente en negar el sentido de cualquier reducción mecánica de lo mental: ni en el pensamiento de Descartes, ni en las mónadas de Leibniz —esos átomos perceptivos—, se puede «entrar como en un molino». El mundo mecánico era aquel del que hablaba Newton, aquel del que pueden predicarse propiedades extensionales, relacionadas, como decía Leibniz, «con las figuras y sus movimientos», pero en ningún modo intencionales: verbos tales como percibir, recordar y pensar quedarían fuera de los límites del mundo mecánico y material. Mentes y máquinas seguían siendo aún entidades radicalmente extrañas en los dos pensadores racionalistas.

Lo que resulta paradójico es que fueran precisamente ellos, Descartes y Leibniz, los que profundizaron en un camino de reflexión (que había transitado antes Llull) que iría a parar finalmente en la negación de la tesis de la diferencia insalvable entre mentes y máquinas: en la Máquina de Turing, los ordenadores y la concepción «mecanicista abstracta» de lo mental, que caracteriza al paradigma dominante de la psicología cognitiva. ¿En qué consistió ese camino? Descartes y Leibniz dieron los primeros pasos hacia el ideal de definir un lenguaje lógico *universal* que fuese capaz de asegurar el rigor deductivo de cualquier clase de razonamiento, y evitar disputas inútiles entre los hombres acerca de todo aquello que pueda resolverse por medio de un *algoritmo.* De forma que los racionalistas ya imaginaron la posibilidad de un *autómata abstracto* y universal, como lo es el de Turing. Lo que no preveían era que esa imagen (sobre todo al encarnarse en la fría piel de los ordenadores) terminaría por echar por tierra su explícita negación de la posibilidad de reducción mecanicista de la mente.

El sueño mecanicista de un lenguaje lógico universal está muy claro en una carta de Descartes a Marsenne de 1629, en la que afirma que es posible concebir un lenguaje en el que «se establezca un orden entre todos los pensamientos que puedan acudir a nuestras mentes, del mismo modo que hay un orden natural entre los números /.../. La invención de tal lenguaje —continúa— es algo que depende de la filosofía verdadera, pues sin ella no es posible confeccionar una relación de todos los pensamientos humanos y ponerlos en orden, ni siquiera separar unos de otros de modo que se tornen claros y simples, lo cual es, en mi opinión, el gran secreto de la obtención de un conocimiento bien fundado. Y si alguno lograse ofrecer una exposición satisfactoria de cuáles sean las ideas simples que hay en las mentes humanas, ideas que constituyen el material de construcción de todos los restantes pensamientos...quedaría casi eliminada la posibilidad de errar» (citado en Nidditch, 1980, pp. 27-28).

Por su parte, Leibniz fue más allá: no sólo imaginó la posibilidad de ese autómata abstracto y universal, compuesto de símbolos y reglas estrictas para su combinación, de un mecanismo que resolvería cualquier problema de razonamiento recurriendo al «cambio e intercambio de símbolos y mediante una especie de Algebra» (Nidditch, op. cit., p. 30), sino que —además de imaginarlo— trató positivamente de desarrollar un lenguaje completo y automático del razonamiento (representando las ideas mediante números que se combinarían de la forma rigurosa y automática que él pretendía). Leibniz no tuvo mucho éxito en ese intento,

que alcanzó forma en el álgebra de Boole, el cual seguía considerando el cálculo lógico como una formulación de *Las leyes del pensamiento* (1854). El intento booleano de dar una fundamentación matemática a la lógica fue la matriz en que se gestó el intento complementario de encontrar una fundamentación lógica completa para la matemática: una pretensión que guió, en todo momento, la elaboración de los *Principia Mathematica* de Whitehead y Russell (1910-1913) y que alcanzó su expresión mas explícita en el llamado «programa de Hilbert»: el intento de definir formalmente, mediante un sistema deductivo *completo* los fundamentos de la matemática, sirviéndose del lenguaje matemático.

Hilbert propuso a los lógicos y matemáticos una tarea gigantesca y muy afín al espíritu con que Descartes trataba de encontrar un lenguaje universal capaz de someter a su rigurosa trama deductiva «todos los pensamientos humanos»: lo que proponía Hilbert era demostrar que el sistema establecido en los *Principia Mathematica* era no sólo coherente, sino también *completo.* Su sueño era la culminación de la vieja pretensión formalista que ha guiado, durante siglos, el desarrollo de la matemática. Comenta Hofstadter que «este objetivo que Hilbert perseguía podrá parecer un tanto exotérico, pero ocupó la cabeza de muchos de los mayores matemáticos del mundo durante los primeros treinta años del presente siglo» (ed. cit., 1987, p. 27). Es sabido que esa pretensión de formalización completa fue demolida por Godel (1931), que demostró que ningún sistema axiomático puede pretender ser *completo* y *coherente* al mismo tiempo (más específicamente, que existen proposiciones verdaderas sobre los números enteros que no pueden demostrarse con la lógica aritmética). En suma, Godel demostró que la pretensión de demostración de completitud de un sistema axiomático, con las herramientas conceptuales del propio sistema, es, por principio, inalcanzable (vid Arbib, ed. esp. de 1987, para una formulación matemática no totalmente incomprensible de] Teorema de la Incompletitud de Godel).

¿No hemos ido demasiado lejos a! referirnos a ciertas consecuencias matemáticas que tuvo a largo plazo la pretensión de un lenguaje universal para el razonamiento? No, porque ese fue justamente el marco intelectual del que nació la Máquina de Turing (1936). Paradójicamente, la definición de una máquina universal tan simple como aquella a la que nos referíamos unas líneas más atrás fue un resultado de un nuevo planteamiento, por parte de Turing de la tesis de la incompletitud de cualquier sistema axiomático formal. La demostración de la universalidad de un mecanismo muy simple, capaz de tratar algoritmos, se originó en la demostración

del carácter *necesariamente* incompleto de todas las lógicas (i.e. de la demostración de que cualquier lógica contendrá enunciados al mismo tiempo verdaderos y no demostrables con el lenguaje de dicha lógica). Turing, Emil Post y Alonzo Church desarrollaron formalismos que demostraban la posibilidad de concebir máquinas abstractas de carácter universal y lo hicieron en relación con la vía de reflexión abierta por el programa formalista de Hilbert y la demostración de su imposibilidad final por parte de Gódel.

Una condición de posibilidad para que se cerrara el círculo de las tres ideas que aparecieron fragmentadas y separadas en el pensamiento de los racionalistas, *máquina, mente* y *lenguaje lógico, completo* y *consistente,* fue despojar al concepto de «máquina» de cualquier clase de connotación material. Con la Máquina de Turing sucede lo mismo que con las mónadas de Leibniz, o la «res cogitans» cartesiana: no es posible entrar en ellas como en un molino, ni su descripción se hace en términos de figuras y movimientos. Es una máquina *formal* y que, precisamente por su *forma* es capaz de imitar a la mente. La forma de la máquina que imagina Turing se compone simplemente de símbolos que responden a una sintaxis completa y bien definida (no importa que tales símbolos sean unos y ceros, proposiciones o sistemas de producciones, por ejemplo, puesto que el cálculo de proposiciones y los sistemas de producciones tienen, en principio, la potencia de una máquina universal de Turing). Por eso decíamos, en otro momento, que la psicología cognitiva, especialmente la que parte de la metáfora del ordenador, se caracteriza por un *mecanicismo abstracto,* y no por un mecanicismo fisicista del tipo de los que sostuvieron Watson o Setchenov.

Uno de los investigadores más profundos del Paradigma ortodoxo C-R (de las computaciones sobre representaciones) en psicología cognitiva, Zenon Pylyshyn, resume muy acertadamente, en pocas palabras, la enorme importancia que ha tenido para nuestra ciencia el enfoque de mecanicismo abstracto, preparado por el pensamiento racionalista y que tuvo su expresión mas clara en la formulación de Turing:

> La obra de Turing puede considerarse el primer estudio de la actividad cognitiva, en que ésta se abstraía en principio por completo tanto de sus fundamentos biológicos como de sus implicaciones fenomenológicas... representa la *emergencia de un nuevo nivel de análisis,* independiente del físico pero de espíritu mecanicista. Hace posible una ciencia de la estructura y la función divorciada de la sustancia material/.../. Dado que habla el lenguaje de las estructuras mentales y los procesos internos, puede dar respuesta a cuestiones planteadas tradicionalmente por los psicólogos (1986, p. 68).

Capítulo 3

LA MAQUINA ABSTRACTA Y LA MENTE REAL. PRIMERAS APORTACIONES A LA CIENCIA COGNITIVA

3.1. Cálculo de proposiciones en el sistema nervioso: el enfoque de McCulloch y Pitts

Hemos visto, en el capítulo anterior, algunas de las consecuencias de aquella travesura que se le ocurrió, a mitad de nuestro siglo, a un matemático genial: Alan Turing. La travesura consistía en imaginar que atribuimos la función mental más eminente y prototípica de las personas, **pensar,** a los ordenadores. Veíamos que uno de los aspectos de la importancia de la demostración de Turing se derivaba de que en ella se define de forma clara, y por vez primera, un nuevo nivel de análisis riguroso y específicamente cognitivo. Un nivel de análisis, como comentaba el científico cognitivo Zenon Pylyshyn, «independiente del físico pero de espíritu mecanicista», que haría posible una ciencia de la mente «de la estructura y la función, divorciada de la sustancia material».

Del mismo modo que es posible conocer el programa que está empleando un ordenador, sin tener en cuenta en absoluto la naturaleza del propio ordenador como sistema físico, así también sería posible conocer los «programas de la mente», sin tener en cuenta la sustancia material en que la mente misma está encarnada. La psicología cognitiva podría entenderse, así, como la ciencia que, mediante procedimientos indirectos, define esas entidades un tanto abstractas y huidizas, pero no menos reales que las materiales, tales como los códigos y lenguajes de la mente, sus algoritmos y procesos, sus representaciones y estrategias. La metáfora de Turing sugería fuertemente la idea de que la mente es, en cierto sentido, una especificación de una máquina de Turing, cuyos íntimos resortes no tendrían por qué ser más misteriosos (aunque quizá sí más complejos) que los de cualquier autómata amental.

La psicología no tendría que preocuparse demasiado del nivel fisiológico de análisis, puesto que su propio plano de análisis sería totalmente **autónomo.** Poco importan, al fin y al cabo, los materiales de que están hechas las «cintas» ideales de las Máquinas de Turing: con independencia de que sean tiras de papel o cintas de algodón, son los unos y ceros que contienen los que permiten realizar cualquier procedimiento efectivo a esa máquina universal.

Sin embargo, esa actitud dualista y la supuesta autonomía de este plano de análisis de los objetos intencionales de conocimiento resultan algo inquietantes desde un punto de vista psicológico, porque los psicólogos no queremos hacer cualquier clase de ciencia cognitiva, sino específicamente *psicología* cognitiva. La demostración deductiva, y completamente abstracta, de que es posible concebir un autómata universal de estados finitos, capaz de computar cualquier función recursiva (i.e. toda función para la que exista un procedimiento efectivo), es muy importante desde el punto de vista de la informática teórica, pero no lo es tanta, para la psicología si se deja en los términos muy abstractos en que Turing la planteó. En realidad, la idealización de la máquina universal implica dejar de lado algunas realidades conocidas y ampliamente aceptadas acerca de los procesos cerebrales y mentales: éstos son limitados en sus recursos de atención y memoria, se producen en unas sustancia compuesta de unidades biológicas diferenciadas (las neuronas) y, a la vez ricamente interconectadas, etc.

Nuestra mente no es ilimitada, como la cinta de la máquina universal de Turing, y ello nos obliga, por ejemplo, a tener en cuenta las limitaciones de memoria que en ella puedan darse. Además, nuestra mente no está implementada en una cinta ilimitada con un dispositivo «lector», sino en un organismo. Mientras que otros científicos cognitivos (los investigadores en Inteligencia Artificial, los lógicos y, en parte, los lingüistas) pueden darse por satisfechos con definiciones abstractas y rigurosas en un plano puramente computacional-representacional, tal como el que definió Turing, los psicólogos tenemos que enfrentarnos a la exigencia adicional de que dicho plano corresponda a una mente real en un organismo real. Ello hace que nuestras posibilidades de aceptación de un formalismo como modelo cognitivo sean mucho más restringidas que las de aquellas ciencias que no tienen por qué sentirse limitadas necesariamente por las características conocidas de las mentes reales.

Por ello, tienen gran importancia histórica los primeros intentos que se hicieron por establecer alguna conexión entre modelos abstractos del

organismo real, y específicamente del sistema nervioso, y los de máquinas lógicas universales, tales como los formulados por Turing, Post y Church. El más influyente de esos intentos fue el realizado por el neurofisiólogo Warren McCulloch y el matemático Walter Pitts, en un artículo de 1943, titulado «Un cálculo lógico inmanente en la actividad nerviosa» (vid. Boden, 1990). En este artículo se daban algunos pasos muy importantes en la dirección de una ciencia cognitiva psicológicamente relevante: (1) se consideraba la lógica como una disciplina esencial para comprender el cerebro y la actividad mental, y (2) se demostraba que el cerebro puede concebirse como un sistema que *incorpora* principios lógicos en sus neuronas y redes neuronales (Varela, 1988). En esencia McCulloch y Pitts (1943, vid Boden, 1990) no sólo argumentaron a favor del principio esencial materialista de que la inteligencia se encarna en el cerebro, sino que probaron rigurosamente que ciertos tipos bien definidos de redes neurales pueden computar determinadas clases de funciones lógicas.

Para realizar esa demostración, McCulloch y Pitts partieron de un modelo idealizado del cerebro, basado en los conocimientos neuro-fisiológicos de su época pero que, al mismo tiempo, eliminaba, por razones de simplicidad, el posible efecto de variables tales como las diferencias temporales en el funcionamiento de las neuronas, las influencias variables de carácter químico y el funcionamiento de las glías. La operación esencial, para su demostración, fue considerar que cada neurona es un autómata-umbral con una o varias entradas binarias y una salida binaria (sobre la base de que la actividad neuronal es un proceso de todo-o-nada). Además, tuvieron en cuenta el artículo de Turing (1936) sobre números computables y las aportaciones de Russell y Whitehead al cálculo proposicional.

Integrando estas ideas, McCulloch y Pitts (1943) probaron varios teoremas acerca de las propiedades lógicas de sus redes neurales idealizadas, como, por ejemplo, las siguientes:

(1) Toda función del cálculo proposicional puede ser realizada por una red neuronal.

(2) Cada red computa una función proposicional que puede ser computada por una máquina de Turing.

(3) Toda función computable por una máquina de Turing puede ser computada por una red neural.

Así, las redes neurales pueden entenderse como autómatas finitos que tienen *potencialmente* (y con la condición de una memoria ilimitada)

el poder universal de cómputo de la lógica de proposiciones o, lo que es lo mismo, de la máquina universal de Turing (vid. McCulloch y Pitts, 1943; Arbib, 1964, para una demostración matemática).

Dado que la máquina universal de Turing tiene una memoria ilimitada, y el cerebro no, la demostración del carácter universal de los cómputos *lógicamente realizables* por redes neurales no bastaba. Era preciso determinar también la configuración de redes específicas capaces de realizar funciones concretas. En los trabajos de McCulloch y Pitts (1943) aparecen numerosos ejemplos: una «neurona abstracta» conectada a otras dos, que posea un umbral tal que sólo se dispare cuando recibe el in-put excitatorio de ambas o cada una de ellas, computa la disyunción no excluyente del cálculo de las proposiciones; si su umbral requiere la suma de las entradas excitatorias de las otras dos para que se dé el disparo, lo que se computa es la conjunción, etc. El trabajo de McCulloch y Pitts abría, de este modo, todo un horizonte de investigación para una psicología cognitiva que definiese el"diseño de las redes neuronales capaces de realizar los mismos cómputos que efectúa la mente humana.

Podemos decir que las concepciones de McCulloch y Pitts sugerían dos vías posibles de investigación:

(1) Una primera alternativa estaría centrada en la idea de *proposición,* y cálculo proposicional, y en una lógica binaria tal como la que define el lenguaje máquina de los ordenadores digitales al uso. Esta línea descuidaría un tanto el análisis del tipo de sistemas (i.e. redes neurales) capaces de realizar los cómputos definidos por aquellos investigadores. En este enfoque, se insiste en la naturaleza simbólica de los lenguajes formales sobre los que se realizan tales cómputos, y los símbolos se consideran como idealizaciones muy abstractas de estados físicos del soporte biológico, cuyo análisis más minucioso se deja completamente de lado. Este primer programa de investigación se fundamenta en una filosofía funcionalista que establece que la mente es al cerebro lo que el programa al ordenador, de tal manera que mentes (y programas) pueden definirse, de forma completa y relevante, sin ninguna necesidad de definir redes neurales (y ordenadores).

(2) La segunda alternativa posible podía consistir en entender que las redes del tipo de las definidas por McCulloch y Pitts (1943) son modelos relevantes (aunque abstractos) de conexiones, umbrales y formas de funcionar del sistema nervioso real. En esta alternativa, se acentuaría la necesidad de estudiar psicológicamente los procesos de cómputo que la mente realiza, pero estableciendo un doble criterio (y no uno solo) con el que valorar los modelos teóricos de la psicología cognitiva: (a) su adecuación a los datos comportamentales y (b) su respeto a principios generales conocidos acerca de la naturaleza y propiedades del sistema nervioso y las redes neuronales.

Es importante destacar que esta segunda alternativa, posible a partir de la línea de trabajo abierta por McCulloch y Pitts, tiene, a primera vista, ventajas importantes en comparación con la primera: (1) no descuida el hecho de que los psicólogos estudiamos funciones mentales reales de organismos reales, y (2) al tener en cuenta, cuando menos en un nivel muy abstracto, propiedades del organismo, resuelve en parte esos problemas de excesiva mediación inferencial, indeterminación y abstracción que tienen los modelos cognitivos cuando sucede, como en la primera alternativa, que éstos se basan solamente en los datos de conducta.

Los dos enfoques que hemos definido pueden identificarse, respectivamente, con el del paradigma clásico C-R (de cómputos sobre representaciones simbólicas en una máquina de carácter digital y esencialmente secuencial), y con la alternativa conexionista (que define las funciones mentales como resultantes del funcionamiento paralelo de redes complejas de «neuronas abstractas»). Es curioso comprobar cómo, a pesar de las aparentes ventajas que hubiera podido tener la segunda opción para los psicólogos cognitivos, fue la primera la que predominó desde finales de los años sesenta hasta mediados de los ochenta, quedando abandonada la alternativa aparentemente más «realista» de las redes neurales. ¿Por qué sucedió esto?

Un factor influyente fue, sin duda, el hecho de que el lenguaje simbólico e intencionalista de la alternativa clásica es mucho más directo, para el psicólogo, que el más molecular y extensional de los modelos de redes neurales. Un segundo factor, al que debemos referirnos ahora, fue el relacionado con la que podemos llamar «la triste historia del perceptrón incomprendido», un título de historia menos fantasioso y absurdo de lo que parece a primera vista.

3.2. **Las limitaciones del perceptrón**

Supongamos una red neuronal que realiza cómputos lógicos, del tipo de los que definieron en su artículo clásico McCulloch y Pitts (1943). ¿Cómo podría aprender? La respuesta es importante para el psicólogo, que sabe que la mente no sólo está «implementada» en el sistema nervioso y tiene cierta capacidad lógica (aunque no suela usarla mucho),

sino también modifica su funcionamiento en virtud de la experiencia. No es una red lógica intemporal, como las abstracciones a que nos referíamos antes, sino un sistema que cambia y *aprende.* Algunas de las primeras investigaciones «pre-conexionistas» sobre redes neurales se dedicaron a este problema: el de explicar su competencia de aprendizaje, una vez que McCulloch y Pitts habían justificado su potencial competencia de deducción lógica.

Entre las aportaciones más influyentes, a este intento de explicar los fenómenos de aprendizaje por redes neurales, están las de Donald Hebb (1949), uno de los precursores más lúcidos del enfoque cognitivo. En su obra de 1949, *La organización de la conducta*, ya apuntaba Hebb al tipo de desarrollos conexionistas que hubieron de esperar más de treinta años para llegar a afianzarse. Al hablar del modelo teórico presentado en ese libro, por ejemplo, señalaba Hebb: «evidentemente, esta teoría es una forma de conexionismo, una variedad de la teoría de la central telefónica, si bien no postula conexiones directas entre senderos aferentes y eferentes. No es una psicología «E-R», si R significa respuesta *muscular»* (p. 21, ed. esp., 1985). Hebb se proponía la enorme tarea de «llenar el hueco» entre los conocimientos psicológicos y neuro-fisiológicos de su tiempo, y para ello recurrió al modelo de redes neurales que aparece en las reflexiones lógicas de McCulloch y Pitts (1943) y en los trabajos neurofisiológicos de Lorente de Nó (1938, 1939, 1943).

En *Organización de la conducta,* Hebb (1949) establecía el principio qiie puede considerarse axioma fundamental de las concepciones del aprendizaje en redes neurales. Es muy simple: cada vez que dos unidades de una red neural, A y B, están simultáneamente excitadas, se incrementa la fuerza de conexión entre ellas. Aunque los detalles del modelo de Hebb puedan haber sido cuestionados posteriormente, éste sigue siendo, en los modelos actuales, el principio del que parten todos los modelos conexionistas del aprendizaje.

Sin embargo, las ideas de Hebb se mantuvieron en un plano especulativo hasta que fue posible construir un modelo real de red capaz de simular el aprendizaje en redes neurales. El primero de estos modelos fue realizado en 1951 por Dean Edmonds y Marvin Minsky, otro importante precursor de la psicología cognitiva. Edmonds y Minsky, con unos cientos de tubos, varios motores y otros cachivache#, hicieron una máquina capaz de aprender algo y que resultaba ser lo suficientemente estimulante como para que Minsky la dedicase su tesis doctoral (Rumelhart, McClelland y PDR Research Group, 1986).

Sin embargo, el trabajo más influyente sobre esas máquinas que eran modelos idealizados del Sistema Nervioso, capaces de aprender, fue el realizado por Franz Rosenblatt, un compañero de Minsky en la Escuela Superior de Ciencia del Bronx. En su libro sobre *Principios de neurodinámica* (1962), Rosenblatt definía las propiedades de ciertos autómatas, dotados para realizar algunas tareas perceptivas y de aprendizaje simples, a los que él denominaba «perceptrones».

En el modelo de Rosenblatt, los perceptrones eran modelos de retinas artificiales asociadas a redes neurales compuestas sólo de dos capas de unidades: unas de in-put y otras de out-put. Eran capaces de aprender a realizar actividades de clasificación de estímulos presentados a su «retina», sobre la base de ciertas propiedades geométricas de los estímulos presentados. Rosenblatt (1962) era muy entusiasta acerca de las posibilidades de los perceptrones, como demuestran algunos comentarios suyos de este tipo: «parece claro que los perceptrones de clase C' introducen una nueva clase de autómatas de procesamiento de la información: por primera vez, tenemos una máquina que es capaz de tener ideas originales» (1959, p. 449). En esencia, Rosenblatt basaba esta suposición en el hecho de que el aprendizaje de tales máquinas estaba basado en un modelo de funcionamiento estadístico, con propiedades emergentes semejantes a las de los fenómenos termodinámicos; un funcionamiento relativamente espontáneo, y auto-correctivo. El aprendizaje de estos sistemas respondía a un principio por el cual, cuando el perceptrón no detectaba un patrón y éste estaba presente, se aumentaban todas las fuerzas de conexión de las unidades activas, a la vez que se disminuía el umbral de las unidades de salida. Y se realizaba la operación inversa cuando el perceptrón señalaba como presente un patrón que no lo estaba.

El entusiasmo de Rosenblatt fue uno de los factores que contribuyeron a que se paralizara la investigación sobre propiedades psicológicas de las redes neurales y de los perceptrones, al demostrarse algunas limitaciones importantes de éstos últimos. En 1969, Minsky y Papert probaron que, aunque los perceptrones puedan servirse de ese «aprendizaje tutorizado por corrección estadística» para aprender eficazmente a clasificar patrones que pueden reconocer, hay otros patrones que sencillamente no pueden ser reconocidos por ellos. Por ejemplo, la disyunción excluyente de la lógica de proposiciones es imposible de «comprender» para un perceptrón compuesto sólo de dos capas de unidades neurales.

La demostración de Minsky y Papert (1969) tuvo un efecto muy desafortunado, que el propio Minsky ha reconocido posteriormente: en los

UNIVERSIDAD DE ANTIOQUIA
BIBLIOTECA CENTRAL

años setenta, se abandonó prácticamente la investigación sobre propiedades computacionales psicológicamente relevantes de las redes neurales. Este abandono se basaba, en realidad, en una confusión: en la suposición de que las limitaciones demostradas para las redes neurales de dos capas podían generalizarse a cualquier clase de redes. Hasta que no se deshizo este equívoco varios años después, no pudieron desarrollarse modelos psicológicos del funcionamiento cognitivo que tuvieran en cuenta la estructura reticular del sistema nervioso..

Esta historia ha tenido consecuencias importantes en psicología cognitiva por una razón en la que debemos detenernos. En el periodo de formulación paradigmática de nuestra ciencia, se daba una situación en que coincidieron, con intervalos temporales de pocos años, las importantes aportaciones lingüísticas de Chomsky —que eran un verdadero modelo ejemplar del paradigma de explicación de la *competencia* lingüística en términos de aplicación de reglas sobre representaciones *simbólicas*—, las explicaciones de Newell, Shaw y Simón (1958) sobre los procesos de solución de problemas, basadas en el supuesto de que la psicología es algo así como la ciencia del «software de la mente en lenguaje simbólico de alto nivel», y finalmente la demostración de Minsky y Papert (1969), según la cual *algunas* redes neurales tenían serias restricciones en sus capacidades de cálculo.

La generalización implícita de esta demostración dejó el camino libre para que predominara una de las dos alternativas, a las que antes nos referíamos: la opción consistente en concebir las funciones mentales como resultantes del cómputo de representaciones simbólicas, en un formato proposicional por ejemplo. La comprensión de ese sistema de cómputo, con el que se identificaba la mente, sería posible sin necesidad de realizar ninguna clase de análisis de las propiedades cognitivas de modelos de redes, semejantes a los perceptrones, y que no parecían ser capaces de explicar actividades mentales que las personas realizamos rutinariamente. Actividades tales como las que permiten comprender enunciados en que se efectúa una disyunción excluyente, de este tipo: «o bien se sigue el modelo conexionista de las redes neurales, o bien se proporciona a la mente un poder computacional suficiente para lo que sabemos de ella, pero no son posibles las dos alternativas a un tiempo».

En suma, el desvanecimiento provisional del interés por las redes neurales dejó el campo libre a una consideración *literal* del reto de Turing: para estudiar la mente no hace falta tener en cuenta su sustrato material; es bastante con el nivel mecanicista y abstracto definido por el

propio Turing; la mente puede entenderse como un mecanismo que computa símbolos; puede imitarse, en sus actividades concretas, mediante máquinas específicas de Turing, y en su conjunto quizá con una máquina potencialmente universal en su capacidad de cómputo, pero limitada por su memoria. La psicología cognitiva puede fundamentarse en un nivel autónomo y abstracto de explicación, independiente del biológico y del social. Al fin y al cabo, los ordenadores, que nos imitan y «piensan», no tienen biología, ni parece que podamos atribuirles un «mundo social» demasiado excitante.

Capítulo 4

CUATRO RESPUESTAS AL DESAFIO DE TURING

4.1. Algunas consecuencias de jugar con Turing

¿Qué consecuencias tiene el hecho de aceptar el desafío de Turing, consistente en suponer que, en un sentido bastante literal, los ordenadores realizan la función «mental» de pensar, y esa función consiste en computar representaciones simbólicas? Dedicaremos este capítulo a realizar una reflexión acerca de esas consecuencias y de las respuestas que se han dado al desafío de Turing. ¿Es la mente un sistema de cómputo?; ¿es, en cierto modo, parecida a una Máquina de Turing?

Antes de realizar esa reflexión, sin embargo, conviene que sinteticemos, en pocas palabras, el argumento que nos ha traído hasta aquí a lo largo de los dos capítulos anteriores. Es, en esencia, el siguiente:

La categoría de los objetos a los que se atribuye mente parece tener la forma de una categoría natural, cuyos límites son imprecisos. Se organiza alrededor de un prototipo central indiscutible, que son las propias personas. Pero sus límites son muy sensibles a variables culturales y evolutivas, así como a concepciones filosóficas explícitas. Los niños y los primitivos, por ejemplo, tienden a atribuir mente a un conjunto más amplio de objetos que los adultos occidentales, y algunos filósofos han hecho coextensiva la mente con todo el ser.

El desarrollo de las ciencias de la naturaleza tiene, en todo caso, la consecuencia de *reducir* los límites de la categoría de los objetos con mente a los organismos vivos, eliminando por completo los residuos animistas de la explicación de los fenómenos físicos. Como consecuencia, las máquinas y fenómenos naturales sin vida se expulsan provisionalmente de la categoría (Fm) —la categoría constituida por los objetos con

mente— del hombre occidental desde la ciencia del Renacimiento. Se establece así una distinción neta entre un mundo cuya descripción puede agotarse en términos extensionales, en lo posible mecanicistas, y un nivel de realidad en el que son aplicables términos intencionales. En el pensamiento racionalista se tematiza explícitamente esa distinción.

El juego de Turing es, por decirlo así, un doble juego: implica volver a desdibujar (por lo menos metafóricamente) las lindes entre las categorías (Fm) y (-Fm) y, al mismo tiempo, aceptar el envite de concebir los predicados Fm en términos estrictamente mecanicistas. Ese juego, que se formula con la pregunta «¿Pueden pensar las máquinas?», tiene sentido: se basa en la demostración previa, por parte de Turing, de que es posible concebir una Máquina Universal (M.U.) capaz de resolver cualquier algoritmo. Tal máquina es de carácter abstracto y formal, y fue ideada en relación con el debate sobre la pretensión formalista de construir sistemas axiomáticos, en matemáticas, que fueran, al mismo tiempo completos y consistentes.

Aunque pueda parecer que la prégunta de Turing, «¿Pueden pensar las máquinas?», es engañosamente animista, su sentido es precisamente el contrario: *des-anima* **la consideración de lo mental, es decir, desliga su análisis del de las propiedades extensionales del organismo; hace imaginar la posibilidad de predicar alguna clase de rasgos (Fm) de sistemas que no están encarnados en un organismo.**

Además el desánimo (esta vez en el sentido más frecuente de la palabra) producido por los primeros análisis sobre las capacidades de cómputo de modelos idealizados del sistema nervioso (redes neurales y perceptrones) no aconsejó, en una primera fase del estudio cognitivo de lo mental, relacionar las explicaciones cognitivas con su marco biológico de origen.

4.1.1. *Dualismo funcionalista*

Este desligamiento entre funciones Fm y organismo nos lleva a detenernos en la consideración de una primera consecuencia de aceptar el juego de Turing. En ella reaparece de nuevo la larga sombra de Descartes, que nos ha acompañado en nuestro camino hasta aquí: se trata de lo que podríamos llamar *dualismo funcionalista,* para distinguirlo del dualismo sustancialista cartesiano. Ha sido este dualismo uno de los supuestos más enraizados y esenciales de la moderna teoría funcionalista de la mente, que sirve de base al paradigma C-R.

El dualismo funcionalista viene a establecer lo siguiente: una Máquina de Turing puede encarnarse en muy distintos tipos de materiales físicos. Lo que importa *para su nivel computacional de descripción* no es la naturaleza y disposición de tales materiales, que es enteramente irrelevante, sino su descripción como sistema simbólico (por ejemplo, mediante una ristra de unos y ceros, o mediante un programa, etc). Del mismo modo, *el nivel mental de descripción,* aun siendo estrictamente mecanicista, puede ser del todo independiente del nivel orgánico (Miller, Galanter y Pribram , 1960; Putnam, 1960; Fodor, 1968; vid. también revisión de Vega y Agapito, 1989). Es una idea que formula con claridad Johnson-Laird (1983):

> La mente puede estudiarse con independencia del cerebro. La psicología (el estudio de los programas) puede hacerse con independencia de la neurofisiología (el estudio de la máquina y del código máquina). El substrato neuro-fisiológico debe proporcionar una base física para los procesos de la mente, pero, con tal de que dicho substrato ofrezca el poder computacional de las funciones recursivas, su naturaleza física no impone restricciones a las pautas de pensamiento (p. 9).

Por su parte, Jerry Fodor es aún más tajante y expresivo en una interesante entrevista reciente con García-Albea (1991):

> Podría ser —dice— que, en cierto sentido, Descartes tuviera razón/.../.Una manera de expresar el punto de vista cartesiano podría ser la siguiente: «Desde una perspectiva científica, no deja de ser una especie de accidente el que los sistemas psicológicos resulten estar encarnados en sistemas biológicos. De hecho, la teoría biológica no te informa mucho acerca de lo que pasa; la que te informa es la teoría de las relaciones funcionales» (p. 10).

La tesis es, en suma, que *existe un nivel mental autónomo, cuya descripción puede realizarse con completa independencia de la descripción del sistema que percibe, piensa, recuerda, etc., como sistema biológico.*

El dualismo funcionalista ha tenido, como mínimo, un efecto positivo en psicología: ha sido decisivo para la diferenciación de ese nivel autónomo, no reductible, de descripción que versa sobre representaciones (conceptos, proposiciones, esquemas, guiones, modelos mentales, estructuras profundas, etc). Junto con la garantía de explicación clara y cientificidad que brinda el carácter mecanicista de las explicaciones computacionales, ha sido un papel decisivo para deshacer las viejas y profundas

aprehensiones de los psicólogos sobre la posibilidad de una psicología científica.

Sin embargo, también hay consecuencias más dudosas: la desvinculación del estudio de los procesos y representaciones mentales de su sustrato neural ha tenido un coste alto, y se basa en presuposiciones que deben cuestionarse.

No está claro, por ejemplo, que el tipo de relación que mantiene un sistema biológico con sus representaciones sea estrictamente comparable al que pueda tener uno artificial con las suyas: ¿por qué, cómo y para qué se hacen conscientes algunas representaciones en el primero y no en el segundo?; ¿hasta qué punto pueden los sistemas artificiales crear representaciones verdaderamente nuevas?; ¿qué condiciones y limitaciones impone la naturaleza del sistema nervioso a las representaciones?; ¿podemos hablar, en rigor, de «representaciones simbólicas» no conscientes, o no serán esas representaciones del psicólogo sobre el sistema, y no de éste sobre el medio?...etc. Los modelos conexionistas, por ejemplo, más sensibles a las propiedades del sistema nervioso que los clásicos, han tenido que abandonar el supuesto de los símbolos no conscientes.

4.1.2. *El carácter simbólico y discreto de las representaciones*

La que acabamos de señalar es una segunda consecuencia de aceptar el juego de Turing en los términos en que éste lo planteó. Las máquinas de Turing, como los programas que se emplean con los ordenadores de tipo von Neumann, consisten en símbolos. Newell y Simon (1976), que están entre los creadores indiscutibles de la psicología cognitiva, plantean esta tesis con la mayor claridad: mentes y ordenadores comparten un aspecto esencial; corresponden a sistemas que son, al mismo tiempo, físicos *y simbólicos.* Son, por así decirlo, máquinas simbólicas implementadas en máquinas físicas. Los sistemas simbólicos físicamente realizables cumplen tres condiciones:

(1) contienen un conjunto de símbolos, que no son, en último término, sino patrones físicos que pueden conjugarse para brindar una estructura (una expresión).

(2) Tienen un conjunto de tales estructuras simbólicas y de procesos con los que operar sobre ellas.

(3) Están situados en un mundo más amplio de objetos a los que designan o interpretan (Newell y Simón, op. cit.).

En realidad, un sistema físico-simbólico es todo aquel en que a ciertos patrones del sistema se les pueden asignar significados arbitrarios y, por medio de un programa, hacer que se comporten de forma coherente con esos significados. Esto (un sistema físico-simbólico) es lo que es cualquier ordenador digital de propósitos generales. Y también lo que son todas las funciones mentales de carácter cognitivo para Newell y Simón: la habilidad de manipular símbolos es, para ellos, la esencia del pensamiento y la inteligencia, como el H_2O lo es del agua, y «la condición necesaria y suficiente para que un sistema físico-simbólico exhiba, en general, acciones inteligentes es que sea un sistema físico-simbólico».

De nuevo, debemos ser cautos, en este caso, en cuanto a lo que significa aceptar las consecuencias del envite de Turing, y de Newell y Simón: implica, en primer lugar, aceptar que las representaciones mentales definidas en el plano computacional —no en el fenoménico— de lo mental, de las que inevitablemente tenemos que dar cuenta los psicólogos cognitivos, tienen un carácter *discreto, arbitrario* y *estructurado* por reglas sintácticas muy estrictas, como lo están los sistemas simbólicos que se emplean con fines de *cálculo* (lógico o matemático). Ello es así hasta tal punto que un sistema simbólico del tipo del lenguaje natural no podría ser tratado (sin una reducción previa a un sistema mucho más inequívoco y estricto) por un sistema físico-simbólico determinista como el que definen Newell y Simón. Pero, ¿son así las representaciones mentales que permiten generar actividades de percepción, memoria, pensamiento y lenguaje?; ¿qué puede querer decir, en realidad, que la oculta mente de las computaciones maneja *símbolos*?, ¿produce símbolos acaso el sistema nervioso? o ¿son —como parece más sensato— formas nuestras de designar sus propiedades como sistema cognitivo?; y, en ese caso, ¿no es excesivamente determinista cualquier modelo físico-simbólico, si se entiende en términos literales, y no como una metáfora o una abstracción de otra cosa?

El supuesto de que la maquinaria cognitiva consiste en símbolos, concebidos como entidades discretas, bien diferenciadas y con reglas semánticas y sintácticas exhaustivamente definidas y precisas, da lugar a dificultades peculiares para comprender, desde el modelo computacional estricto, *fenómenos* tales como las imágenes mentales. En este sentido, Pylyshyn (1984) es muy claro:

> la base de toda la comprensión formal —dice— se encuentra en la noción de símbolo atómico discreto. En realidad constituye la base de todos los sistemas

de pensamiento, expresión o cálculo para los que disponemos de una *notación.* Resulta importante destacar que esta idea tiene profundas raíces en lo que a veces se ha llamado la tradición intelectualista, pero nadie ha conseguido definir otro tipo de átomo a partir del cual pueda derivarse la comprensión formal. Por tanto, no ha de extrañar demasiado que muchos de nosotros seamos reacios a abandonar esta fundamentación de la ciencia cognitiva, a pesar de las frecuentes exhortaciones que recibimos para que aceptemos símbolos con propiedades intrínsecas tan diversas como las propiedades continuas o análogas /.../. El problema consiste en que las nociones de este tipo carecen de fundamentación sistemática. No sabemos qué podemos hacer con ellas» (ed. esp. de 1988, p. 8).

4.1.3. *Formalismo y sistemas guiados por su sintaxis*

El comentario de Pylyshyn, que acabamos de recoger, es muy revelador del tipo de explicaciones a que dirige la aceptación literal del desafío de Turing: se trata, ante todo, de explicaciones *formales.* En psicología cognitiva en general, y en el paradigma C-R sobre todo, explicar un conjunto de fenómenos equivale a *definir una pauta formal capaz de generar ese conjunto de fenómenos.* Se supone que esa pauta es equivalente, en el sentido fuerte (Fodor, 1968) a lo que sucede en la mente. Ese estilo explicativo ha tenido consecuencias positivas muy importantes, al permitir alcanzar niveles de rigor y precisión en la explicación de los fenómenos mentales y comportamentales que no eran posibles antes del desarrollo de las explicaciones cognitivas.

Sin embargo, también en este caso es necesario comentar brevemente los efectos más dudosos de la propensión formalista: uno muy evidente es que se ha producido históricamente la tendencia a estudiar, sobre todo, conjuntos de fenómenos fácilmente formalizables (la concentración de estudios de razonamiento en silogismos, o el predominio de investigaciones sobre aspectos sintácticos en la primera década de desarrollo de la psicolingüística, son ejemplos claros). Pero hay un aspecto más importante y profundo que ése: una máquina formal, del tipo de una máquina de Turing o un programa de ordenador, se guía en su procesamiento exclusivamente *por su sintaxis.* No actúa, en realidad, en función de los *contenidos semánticos,* de los significados de sus estructuras «de conocimiento» sino en virtud de la pura *forma* de las representaciones. Aun cuando a los símbolos arbitrarios del sistema se les asignen, como dicen Newell y Simón (1976), funciones designativas o interpretativas, seguirá siendo la forma de éstas, y no su contenido, la que dirija los procesos.

Los defensores más lúcidos del paradigma C-R son perfectamente conscientes de la dificultad que aquí se plantea. Así Fodor señala en *Psychosemantics* (1988) lo siguiente: «éste es, a grandes rasgos, el proceso que se supone que sigue el asunto: las propiedades causales de los símbolos se conectan con sus propiedades semánticas, por la *vía de la sintaxis.* La sintaxis de un símbolo es una de sus propiedades físicas de orden superior. En una primera aproximación metafórica, podemos entender la estructura sintáctica de un símbolo como una caracterización abstracta de su forma. Debido a que, a todos los efectos y propósitos, la sintaxis se reduce a la forma, y debido a que la forma del símbolo es un determinante potencial de su papel causal, es fácil ver cómo puede haber medios en que el papel causal de un símbolo correlacione con su sintaxis /.../ . Ahora bien, sabemos por la lógica moderna que ciertas relaciones semánticas entre los símbolos pueden ser, como si dijéramos, «emuladas» por sus relaciones sintácticas/.../ Los ordenadores dan una solución al problema de mediar entre los poderes causales de los símbolos y sus propiedades semánticas» (pp. 18-19). Pero, ¿puede ser completa esa reducción sintáctica de la semántica?

4.1.4. *Flexibilidad mental y sistemas rígidos*

El problema que plantea la pregunta anterior es importante. Un sistema simbólico de cómputo que se guíe finalmente sólo por las relaciones sintácticas preestablecidas entre sus símbolos (y, en último término, *también* las relaciones semánticas serían sintácticas, en esta perspectiva, como sugiere Fodor en su comentario anterior) es, al mismo tiempo, *universal* y *rígido.* «Rígido» en el sentido de que es absolutamente determinista: sus reglas y representaciones no se conectan, en la mayoría de las concepciones formuladas hasta ahora, por relaciones probabilísimas sino que se desencadenan mutuamente de forma inflexible. Es posible que no tenga por qué ser así, pero resulta difícil concebir un sistema que se componga de una sintaxis compuesta de estructuras de símbolos atómicos, con poderes causales de cómputo, y que admita enunciados probabilísticos. Como veremos en otro momento, los sistemas conexionistas son más flexibles, pero no sería exacto decir de ellos que «se guían por la sintaxis».

¿Puede concebirse la mente como un sistema determinista rígido? No parece probable que un sistema de ese tipo, aun siendo finalmente «uni-

versal», pueda haber resultado de un largo proceso evolutivo, en que la capacidad de adaptación flexible al medio, y la presencia de variaciones *continuas,* no discretas y cualitativas, tuvieron que jugar papeles importantes. En términos generales, nuestros modelos actuales de la mente son excesivamente rígidos. Su dificultad para dar cuenta de las variaciones individuales entre sujetos tienen mucho que ver con esa característica (vid. Juan Espinosa y Colom, 1989; Colom y Juan Espinosa en prensa). No hace falta recordar que sin variación no hay evolución, pero sucede que en un sistema simbólico rígido, como el que propone la metáfora clásica del ordenador, toda variación es cualitativa y potencialmente catastrófica (i.e. hace que no se compute «lo mismo de otra manera», sino «otra cosa»). La opción de incluir rutinas de aleatorización, asignar pesos probabilísticos a producciones o proposiciones (como hace, por ejemplo, Anderson, 1976, 1983) no termina de resolver el problema: consiste en adaptar *post hoc* un sistema rígido, en vez de comprender —como hacen los conexionistas— que la rigidez es una propiedad de un cierto plano del discurso sobre la mente, y no una propiedad de la propia mente (i.e. un resultado de la consideración *metafórica* de la mente computacional como simbólica).

4.1.5. *La irrelevancia computacional de la conciencia*

La inflexibilidad determinista de los autómatas simbólicos, que son derivados más o menos directos de las máquinas de Turing, Post y Church, se relaciona muy directamente con la última consecuencia de jugar con Turing que vamos a comentar: desde el plano computacional de explicación, no ha resultado hasta ahora fácil comprender el papel que pueda tener la conciencia, el hecho de que el sistema se presente fenómenos a sí mismo y, en cierto modo, se contenga a sí mismo. Esas son propiedades que no han necesitado las máquinas de Turing para alcanzar ese ideal de «razón universal», al que ya nos hemos referido. El propio Turing (1950) preveía el problema en el artículo que originó la metáfora del ordenador, y daba una hábil respuesta que, sin embargo, no lo deja solucionado:

Argumento de la conciencia: /.../ Sólo cuando una máquina sea capaz de escribir un soneto o componer un concierto por haber experimentado pensamientos y emociones, y no por una conjunción casual de símbolos, admitiremos que pueda ser igual al cerebro es decir, que no solamente escriba, sino que conozca que escribe/.../

Este argumento parece ser una negación de lá validez de nuestro *test.* Según la formulación más extrema de tal punto de vista, la única manera por la cual uno podría estar seguro de que una máquina piensa, consistiría en *ser* la máquina y sentirse pensar uno mismo /.../. En resumen, pues, creo que a la mayoría de los que sostienen el argumento de la conciencia se les podría persuadir de que lo abandonasen, antes de verse forzados a caer en una posición solipsista /.../.

No quisiera dar la impresión de que pienso que no hay misterio alguno con respecto a la conciencia. Hay, por ejemplo, algo de paradójico en lo que se refiere a cualquier intento de localizarla. Pero no creo que sea necesario solucionar estos misterios antes de poder responder a la pregunta que nos ocupa en el presente artículo (ed. esp. de 1985, pp. 37-40).

Sin embargo, el astuto encierro de la conciencia en el reducto solipsista e intrasubjetivo del que tanto la cuesta salir a la pobre, no responde a algunas preguntas importantes: dado que el hombre es capaz de concebir sistemas universales de cómputo que no requieren de esa prisionera oculta del solipsismo, ¿para qué la ha desarrollado la naturaleza? Si cumple —como la gente intuye— un papel fundamental, ¿cabe una descripción completa de la mente como sistema de cómputo? Quizá todo el malentendido se base en otro equívoco de fondo: ¿No será que los ordenadores, en cierto modo, sí tienen alguna clase de conciencia, y que del mismo modo que el ordenador es una metáfora de la mente como sistema de computación, la mente del que pone y quita programas, los inventa y desarrolla, es una especie de metáfora de la conciencia del ordenador?

El complicado problema de relaciones entre la mente-uno (conciencia) y la mente-dos (cómputos), a que ya nos referíamos en el capítulo anterior, está ciertamente lleno de equívocos, y hay que reconocer que no toda la culpa es de la segunda. Los equívocos empiezan en la propia definición de lo que es la conciencia: decir sólo que es un sistema que *presenta* al organismo como un todo resultados finales de los cómputos en forma de fenómenos que guían la conducta, es decir muy poco. Hasta la segunda mitad de los años ochenta, los psicólogos cognitivos más cercanos al paradigma C-R se ocuparon muy poco de la conciencia. Se han dado pasos importantes después, en obras de síntesis, como las de Baars (1988), Jackendoff (1987) y Johnson-Laird (1988). El primero de ellos, por ejemplo, emplea términos del vocabulario de procesamiento para resumir las funciones de la conciencia: «debugging», «flagging», «edi-

ting», «priorizing and access-control», «meta-cognitive or self-monitoring function», «auto-programing and self-maintenance function», etc. Es muy difícil despejar de confusión la definición de un concepto tan gravosamente cargado de funciones como las que aparecen en esos términos del vocabulario C-R queemplea Baars.

Johnson-Laird (1988), por su parte, dice que «la conciencia simple —sencillamente, darse cuenta de sucesos como el dolor— puede tener sus orígenes en la aparición, a partir de una tela de araña de procesadores paralelos, de un monitor de alto nivel» (p. 336) y la conciencia compleja y auto-reflexiva podría entenderse como una máquina de Turing «con poder de auto-descripción» (p. 340), pero él mismo reconoce el carácter muy especulativo y la falta de formalización y apoyo empírico que pueden tener por ahora estas ideas.

En suma, aceptar el juego de Turing es aceptar un juego peligroso en algunos aspectos. Sus reglas son claras: la mente puede entenderse como un sistema de cómputo sobre símbolos discretos. Ese sistema tiene una misteriosa propiedad de ser consciente, propiedad que no termina de encajarse en su descripción como tal sistema de cómputo. La mente, como sistema simbólico puede entenderse en términos de lo que hemos llamado «determinismo rígido» y, entonces, resulta difícil explicar su flexibilidad adaptativa y el hecho, difícilmente controvertible, de que parecen existir diferencias individuales entre los elementos de una misma especie de los que se hace predicados (Fm) —si bien hay que reconocer que las diferencias individuales entre los ordenadores quizá sean más desdeñables—. Finalmente, el nivel de descripción de la mente como «mente-dos» es independiente por completo de la descripción del sistema que la soporta como maquinaria orgánica o artificial. Pero, en ese caso, se plantea el problema de la tendencia a desvincular a la mente de su pasado filogenético y de las condiciones orgánicas que la soportan. ¿No es demasiado para un juego?

4.2. Cuatro respuestas al desafío de Turing

En la psicología cognitiva y la filosofía de la mente, se han dado cuatro respuestas principales (con numerosas variantes de detalle, en las que no entraremos) al juego con el que nos desafía Turing —considerar que los ordenadores son objetos a los que puede atribuirse la función de pensar— y a la pregunta que acabamos de hacernos al final del capítulo anterior. La pregunta era si las consecuencias de aceptar las premisas de Turing no son excesivas para constituir una ciencia objetiva de la mente.

Antes de analizar esas respuestas, recordemos los términos en que hemos planteado nosotros el juego: consiste en imaginar qué es lo que sucede cuando se desdibuja la distinción entre las categorías (Fm) y (-Fm) que habían sido demarcadas desde la ciencia mecanicista del Renacimiento; implica, así, tratar de desarrollar una concepción rigurosa y mecanicista de lo mental, basada en dos supuestos complementarios: en un sentido preciso, y dejando de lado esas «perchas para colgar el misticismo» que decía Watson (1930), ciertas máquinas piensan, y —la otra cara de la moneda— las funciones Fm pueden entenderse como máquinas abstractas. Además, el tipo de máquina imaginado por Turing se compone de símbolos discretos, con una sintaxis determinista y bien definida, y su descripción es independiente de la del soporte físico en que se implementa, con tal de que éste asegure el poder computacional necesario (por ejemplo, el poder de computar cualquier «función recursiva»; vid. Arbib, 1964, para una definición matemática de las funciones recursivas).

Las cuatro respuestas fundamentales al reto de Turing pueden resumirse así:

(1) No aceptar, de ningún modo, la alternativa de desdibujar la categorización tradicional (Fm) (-Fm). Implica hacer el juego de Turing, pero responder escuetamente a su pregunta, «¿pueden pensar las máquinas»: **No,** no pueden pensar. Pensar es una propiedad (privada) de un organismo; por consiguiente, dejemos las cosas como estaban: (Fm) y (-Fm). Ni las máquinas piensan, ni las mentes son máquinas.

(2) *Aceptar, con todas sus consecuencias, el desafío de Turing: hay un sentido literal en que la categoría (Fm) incluye mentes y máquinas.* También en este caso se acepta el juego, pero con tanta seriedad que deja de ser un juego. La respuesta al desafío de Turing es, en esta posición, igual de clara y lacónica: Sí, pueden «pensar», si es que «pensar» es algo preciso. **Las mentes son precisamente máquinas simbólicas.** La eliminación de la distinción tradicional (Fm) y (-Fm) entre todas las mentes y algunas máquinas es la única alternativa coherente para explicar y comprender las mentes.

(3) *Aceptar el juego como una metáfora.* Es ésta la alternativa, menos radical que las dos anteriores, que han adoptado —implícita o explícitamente— los psicólogos cognitivos en su gran mayoría. Puede resumirse

así: «¡hombre!...no se ponga usted tan serio. Ya sabemos que ni las mentes son, en último término, máquinas, ni las máquinas mentes. Pero el *juego* es útil como *metáfora*...no hay que extremar las cosas. Por consiguiente, aceptemos el juego de emplear el vocabulario del *procesamiento de la información*, sin comprometernos con esas consecuencias tan negativas que nos amenazaban, como nubarrones, en el apartado anterior». En esta opción se «juega» a desdibujar los límites clásicos (Fm)/(-Fm), sin llegar a la seria conversión del juego en algo más, que caracterizaba a la posición segunda.

(4) *Limitar las reglas del juego.* Es la opción más reciente. Implica una respuesta muy matizada: aceptando que la mente es un sistema de cómputo, no es el tipo de máquina simbólica que describe Turing. En realidad, la metáfora de la mecánica no es tan exacta como puedan ser otras metáforas tales como las termodinámicas, por ejemplo. Si se tienen en cuenta las propiedades de los sistemas que tienen mente —sistemas nerviosos— se ve que tales sistemas son capaces de cómputos complejos (quizá no tan universales como los de la máquina de Turing, ¿pero acaso es realmente universal la razón humana?) sin necesidad de acudir a supuestos como el de los símbolos inconscientes, que no son más que metáforas de propiedades emergentes del funcionamiento de tales sistemas.

En este capítulo analizaremos algo más a fondo las cuatro alternativas descritas, empezando por las dos más extremas, (1) y (2) que implican una respuesta lacónica, afirmativa o negativa, a la pregunta de Turing con todas sus implicaciones. Después examinaremos las dos respuestas más moderadas, (3) y (4).

4.3. **La mente de la habitación china** y **el martirio de la psicología cognitiva**

La primera respuesta niega todo sentido a la disolución de las categorías tradicionales (Fm)/(-Fm) que implica la apuesta de Turing. Es una respuesta que probablemente tenga muchos seguidores «implícitos», pero que ha sido argumentada, de forma más sistemática, por algunos filósofos de la mente, como Hubert Dreyfus (1979) y John Searle (1980), e incluso, hasta cierto punto, por algunos teóricos de la Inteligencia Artificial, como Fernando Flores y Terry Winograd (1989). Esta posición no

es defendida por ningún psicólogo cognitivo y, si se extrema, termina, como veremos enseguida, en la *negación de la posibilidad misma de la psicología cognitiva.* Para resumir brevemente este primer enfoque, me referiré, sobre todo a los argumentos más claros y contundentes: los de John Searle (1980. 1983).

Searle (1980, cit. reed. 1990) hace una finta muy ingeniosa: comienza por aceptar, en sus términos el juego de Turing. Hagámoslo también nosotros para comprender su posición. Supongamos, entonces, que tenemos una máquina universal de Turing —instanciada en un ordenador— con un programa especialmente adecuado para hacer el test de pensamiento: se trata del diseñado por Schank y Abelson (1977) para simular procesos complejos de comprensión. Es éste un buen ejemplo, porque la explicación de los procesos de comprensión dada por Schank y Abelson ha sido una de las más refinadas e influyentes en el estudio de esos procesos en psicología cognitiva.

El programa de Schank y Abelson que, como es bien sabido, posee «scripts» o guiones (es decir, esquemas de alto nivel que permiten organizar y anticipar información sobre sucesos que se dan en ciertos marcos), tiene un guión que se ha utilizado como ejemplo prototípico: el de restaurante. Como además posee «estructuras conceptuales» abstractas, complejas y precisas, pasa muy bien el test de Turing cuando se le hacen preguntas sobre restaurantes. Produce una clara impresión de inteligencia. Si se le cuenta una historia como ésta «Un hombre fue a un restaurante y pidió una hamburguesa. Se la pusieron quemada. Se fue furioso y sin pagar» y se le pregunta: «¿se comió la hamburguesa?», el programa contesta: «No, no se la comió». Si la historia empieza igual y termina: «el hombre se fue tan contento, y dejó una fuerte propina, ¿se comió la hamburguesa», responde: «Sí, se la comió». La cuestión es la que ya se planteaba Turing: ¿podemos decir que el programa *comprende* lo que se le narra y pregunta?. La respuesta de Searle (op. cit.) es, sin embargo, la contraria a la de Turing: *no*, no podemos decirlo de ningún modo. ¿Por qué?

Para ilustrar por qué no es posible decir que el programa comprende, siguiendo el mismo juego de Turing, Searle se imagina un experimento «hecho con el pensamiento», un *Gedankenexperiment*, que consiste en lo siguiente: él mismo está encerrado en una habitación —como lo estaban el hombre y la máquina de Turing— y fuera de ella hay unos interrogadores bilingües, que hablan y comprenden inglés y chino. Empieza Searle por confesar que él mismo no sabe una palabra de chino.

En la habitación hay una gran pila de tarjetas con caracteres chinos, que no comprende en absoluto. Desde fuera, le pasan otro conjunto de símbolos chinos junto con un libro de instrucciones muy simples que permiten asociar los primeros caracteres a los segundos. Por razones insondables para Searle, los bilingües de fuera llaman al primer conjunto de caracteres «historia», al segundo «guión», y al manual de instrucciones «programa».. El juego consiste, esta vez, en que tales interrogadores pasan unas veces sus historias y preguntas en chino, y otras veces en inglés. Desde el punto de vista de los que preguntan desde el exterior, el Searle oculto de dentro de la habitación, que sigue obedientemente lo que le indica el libro de instrucciones, «comprende chino», y pasa con éxito el test de Turing. Sus respuestas son tan buenas en chino como en inglés. Pero los «psicólogos objetivistas» de fuera se engañan: Searle decididamente *no* comprende nada de chino.

Del mismo modo que el ordenador era una metáfora de la mente en el test de Turing, Searle es una metáfora del ordenador en el test de Searle: «en lo que se refiere al chino —dice— yo me comporto simplemente como un ordenador; realizo operaciones de cómputo sobre elementos formalmente especificados. A propósito del chino, soy simplemente una concreción de un programa de ordenador» (ed. 1990, p. 69): una especie de metáfora de los guiones de Schank y Abelson. AI volver del revés el test de Turing, y convertir al Searle de la habitación en metáfora del ordenador, el verdadero Searle propone un argumento de gran fuerza expresiva para oponerse a las pretensiones explicativas del paradigma clásico C-R. El problema fundamental es que este paradigma presupone que dar cuenta de procesos tales como pensar, comprender, o recordar consiste en especificar un conjunto de operaciones de cómputo sobre símbolos, de tal manera que las operaciones estén determinadas por la *forma* de las expresiones simbólicas (de esto ya hemos hablado) y todo ese programa explicativo implica una profunda confusión para Searle.

En tanto en cuanto el programa se define en términos de operaciones de cómputo sobre elementos definidos en términos puramente formales —dice Searle—, lo que el ejemplo sugiere es que éstos por sí mismos no tienen ninguna conexión interesante con la comprensión. Ciertamente no son condiciones suficientes de ella; y nadie ha dado razones de peso para suponer que sean condiciones necesarias o, ni siquiera, que hagan una contribución significativa al proceso de comprensión (pp. 70-71).

La evaporación de los límites (Fm)/(-Fm) para mentes/máquinas *formales,* sugerida por Turing no tiene razón de ser: a veces decimos que «la puerta *sabe* que tiene que abrirse, porque tiene una célula fotoeléctrica.», por ejemplo. Ese lenguaje animista es interesante en sí mismo, en tanto que expresa una tendencia del pensamiento humano a ampliar metafóricamente el ámbito de lo intencional. Pero no es más exacto ni más relevante en el caso del ordenador que en el de la puerta. Las puertas no piensan ni comprenden, los ordenadores tampoco. Las mentes sí.

La razón de fondo de Searle es que la mente es un sistema *intencional* y los ordenadores no lo son: «el punto principal del presente argumento —afirma— es que ningún modelo puramente formal será nunca suficiente por sí mismo para producir intencionalidad, porque las propiedades formales no son, por sí mismas, constitutivas de intencionalidad, ni tienen por sí mismas ningún poder causal, excepto la capacidad de producir, cuando son instanciadas, el siguiente estadio del formalismo si la máquina está funcionando» (p. 82). Esa misteriosa propiedad de intencionalidad, de *ser acerca de algo,* que Searle afirma de la mente y niega de los ordenadores, no es un resultado de las características formales que pueda tener lo que la mente hace, sino precisamente de la *materia* de que está hecha: tejido nervioso. Aquí es donde Searle pasa de su proposición negativa (las máquinas formales no piensan ni comprenden, en tanto que ello se haga depender de sus propiedades formales) a una proposición positiva mucho más arriesgada y, en el fondo, difícil de aceptar y comprender: **la** intencionalidad, **que es el** rasgo **esencial** de lo **mental,** es **una** propiedad **de la materia específica de la que está hecha la mente, de la materia** biológica. Veamos cómo, en la forma en que la plantea Searle, esta tesis positiva terminaría con la pretensión cognitiva de una ciencia objetiva de la mente.

Para Searle, la enorme inversión de pensamiento que ha supuesto el aceptar —radical o moderadamente— el reto de Turing, en las ciencias cognitivas, es una inversión a fondo perdido, y ello es así porque tal inversión se basa en tres supuestos que no son, en modo alguno, justificables:

(1) La confusión entre forma y contenido: los estados mentales tienen una característica esencial de intencionalidad, que se define justamente como posesión de un contenido, no de una forma. La creencia de que «los patos son ovíparos» no consiste en la definición morfosintáctica de esa oración y, de hecho, sigue teniéndose cuando se formula con ora-

dones diferentes, tales como «las hembras de pato ponen huevos». La creencia es la misma porque el significado de las dos oraciones es el mismo. Podemos resumir esta primera observación así: los estados mentales tienen, como propiedad del máximo interés, el hecho de que *significan* algo. Los símbolos del ordenador, los cómputos que efectúa, no significan nada para él. En realidad, sus símbolos no son tales, puesto que carecen de una de las dos caras que definen a un símbolo: el significado.

(2) **Conductismo residual de la Inteligencia Artificial, o carácter falso del test de simulación de Turing:** «En gran parte de la I.A. hay una especie de operacionalismo o conductismo residual»: ésa es la segunda confusión. El hecho de que puedan hacerse idénticas las «entradas» y «salidas» de ordenadores y personas (como en el «inteligente» programa de comprensión de Schank y Abelson) no quiere decir que pueda decirse, en un sentido relevante, que los primeros posean nada que se parezca a la mente (intencionalidad). «Mi calculadora de mesa tiene capacidades de cálculo, pero no intencionalidad... El test de Turing es típico de una tradición acrítica conductista y operacionalista, y creo que si los investigadores de la I.A. repudiasen por completo el conductismo y el operacionalismo, se eliminaría gran parte de la confusión entre simulación y duplicación» (pp. 85-86).

(3) **Dualismo:** El tercer supuesto resulta sorprendente, dado el segundo. Resulta que el operacionalismo residual de la I.A. se asocia a un dualismo residual: en lo que a la mente concierne, la materia no importa, y «en lo que a la I.A. concierne, el mismo programa podría ser realizado por una máquina electrónica, una sustancia mental cartesiana, o un espíritu universal hegeliano» (p. 86). Ya hemos visto, en el apartado anterior, que ese dualismo residual ha sido uno de los principios esenciales del enfoque funcionalista, en que se ha fundamentado filosóficamente todo el proyecto de la ciencia cognitiva (incluyendo a gran parte de la psicología cognitiva).

Bien. Imaginemos, entonces, que aceptamos esas críticas de Searle: abjuramos del formalismo, el conductismo residual y las sospechosas propensiones dualistas de la I.A. Supongamos también que hacemos nuestra una tesis de vieja tradición psicológica (Brentano, 1874), según la cual lo que define más claramente a la mente es la intencionalidad, el ser acerca de algo, el tener contenidos. ¿Cómo explicar esa propiedad esencial de la mente sin caer en los supuestos que hemos abandonado? La intrigante respuesta de Searle es ésta: sucede que tal propiedad pertenece a la materia biológica, y sólo a ella. La afirmación de Searle es clara en este sentido:.

> ¿Puede pensar una máquina? Mi punto de vista es que *sólo* una máquina puede pensar, y en realidad sólo un tipo especial de máquinas, a saber, los cerebros y las máquinas que tienen los mismos poderes causales que los cerebros. Y que ésta es la razón principal por la que la I.A ha tenido poco que decirnos acerca del cerebro: porque no tiene mucho que decir tampoco acerca de las máquinas. Por su propia definición, trata sobre los programas, y los programas no son máquinas. Con independencia de lo que pueda ser la intencionalidad, es un fenómeno biológico, y *es tan probable que dependa causalmente de la bioquímica específica de sus orígenes como la lactación, la fotosíntesis y otros fenómenos biológicos* (86-87).

Del mismo modo que nos hemos detenido a considerar, en otro momento, las implicaciones y consecuencias que podría tener aceptar la tesis de Turing tal como él la planteaba (y mas explícitamente Newell y Simón, en sus escritos sobre el sistema físico-simbólico, vid. 1976), debemos, para ser justos, reflexionar sobre las consecuencias últimas de la tesis de Searle. Son mucho más graves para los que tratamos de investigar en psicología cognitiva. El propio Searle (1983) las establece de forma muy explícita: parece haber un vacío insalvable entre las explicaciones de sentido común de la conducta (en términos intencionales) y las explicaciones neurofisiológicas, «y alguno de los mayores esfuerzos intelectuales del siglo XX han consistido en intentar llenar ese vacío, en lograr una ciencia de la conducta humana que no fuera solamente la psicología de sentido común de la abuelita, pero que no fuera tampoco neurofisiología científica. Hasta la época presente, los esfuerzos para llenar el vacío han sido, sin excepción, fracasos /.../los esfuerzos por llenar el vacío fracasan porque no hay ningún vacío que llenar» (pp. 49-50). Y añade, en otro momento, que las oportunidades de éxito del cognitivismo «son casi tan grandes como las oportunidades de éxito del conductismo hace quince años. Es decir, ...prácticamente cero» (p. 63).

La conclusión es completamente coherente con las premisas de Searle: el nivel intencional de descripción de lo mental, que empleamos en nuestras interacciones cotidianas, no es más que un nivel más molar de referirse a propiedades del sistema nervioso que éste tiene como sistema material, por ejemplo, en un plano bioquímico de explicación (el cerebro hace mente intencional de forma similar a como las plantas hacen fotosíntesis). No hay vacío entre uno y otro plano explicativo: la psicología

cotidiana del sentido común es sólo la forma holista de definir lo mismo que define la neurofisiología y, en último término, la bioquímica: «además de un nivel de estados mentales, como las creencias y deseos, y un nivel de neurofisiología, no se necesita nada que rellene el hueco entre la mente y el cerebro, porque no hay hueco que llenar» (p. 64).

El diagnóstico de Searle es pesimista: para él, las pretensiones de la psicología cognitiva se basan en una gigantesca confusión conceptual. Si no hay un *plano explicativo* y *autónomo* entre las explicaciones extensionales de los neurofisiólogos y lo que llamaba en otra parte «sujeto de atribución de la psicología natural» (Rivière, 1987), los psicólogos cognitivos tenemos solamente dos alternativas: la más radical —y, en el fondo la más sensata— es dejar la profesión; la de compromiso, desempolvar rápidamente los libros de James y Wundt, e intentarlo de nuevo con la psicología introspectiva.

Afortunadamente, creo que quien tiene, en este caso, una gran confusión conceptual es el propio Searle, tan lúcido en muchas otras ocasiones. Las razones que ofrece implican una mezcla inconsistente de planos explicativos, y el pronóstico de la psicología cognitiva es algo menos sombrío de lo que él cree. Veamos por qué.

Las personas no tenemos sólo ideas implícitas, profundamente enraizadas, sobre lo mental, sino también sobre el mundo físico (vid., por ejemplo, Pozo, 1987). Ya hemos comentado que tales ideas están, por ejemplo, infiltradas de propensiones animistas. Tienen que ser modificadas profundamente en la enseñanza, por esa y otras razones. Para la gente no es obvio el principio de conservación del movimiento; nuestra física intuitiva tiende, más bien, a considerar que el «reposo» es el estado natural de los cuerpos. Además, a medida que las explicaciones físicas se alejan de ese nivel «mesocósmico», de dimensiones intermedias, del que hablan Delbruck (1986) y Vóllmer (19S4), resultan menos intuitivas. Supongamos que nos situamos en ese plano, muy poco intuitivo, del «microcosmos» de la física cuántica y la física de partículas elementales. La tesis de Searle, trasplantada al mundo físico, vendría a equivaler a la siguiente —y sorprendente— afirmación: «Además de un nivel de física intuitiva, sobre los objetos que vemos y tocamos, y un nivel de física cuántica, no se necesita nada que rellene el hueco entre las "cosas" y los "cuantos", porque no hay hueco que llenar».

Esa afirmación equivaldría a realizar, para el plano físico, exactamente las mismas operaciones que Searle considera justificables en el psicológico: (1) presuponer que las teorías cotidianas implícitas son modelos

molares de las explicaciones cuánticas (es obvio que no lo son) y (2) eliminar, de golpe, los planos de explicación en que aparecen propiedades que merecen ser estudiadas por sí mismas, y tienen *leyes propias*, en tanto que emergen del funcionamiento global de sistemas: sistemas atómicos, mecánicos, termodinámicos, etc. No hay ninguna razón seria para suponer que no sea necesario admitir la existencia de un plano cognitivo, con leyes propias y autonomía como tal sistema, que facilite esa larga trayectoria entre la intención y la bioquímica de la sinapsis, que Searle (1985) pretende hacer en dos pasos.

Esa admisión no tiene por qué hacernos caer en los brazos del dualismo: lo que se plantea es qué es más relevante para la explicación de la capacidad de las personas y animales de poseer estados intencionales (i.e. mente), si la descripción bioquímica de lo que sucede en las sinapsis o, por ejemplo, la descripción de propiedades que tiene el sistema nervioso, *como sistema de cómputo*, resultante del funcionamiento masivo y molar de millones de neuronas. Por lo que parece, por ahora, para definir tales propiedades (vid. McClelland, Rummelhart y PDP Research Group, 1986) no ha sido necesario penetrar en las íntimas relaciones bioquímicas que se mantienen en los alrededores de las membranas sinápticas. De igual manera que, para describir rigurosamente propiedades termodinámicas relacionadas con la Ley de Entropía, no resulta indispensable (y sería engorroso) hablar de las aún más íntimas y minuciosas relaciones entre los cuantos.

El progreso científico implica frecuentemente dos pasos: (1) la diferenciación de niveles explicativos, en que pueden formularse leyes que caracterizan el funcionamiento de los sistemas cuando alcanzan ciertos grados de organización, y (2) el establecimiento de relaciones entre los niveles previamente diferenciados. No hay ninguna razón para suponer que este doble proceso, que tan buenos resultados ha dado en física y biología, no sea necesario también en psicología. Entre la bioquímica cerebral y la psicología intencionalista cotidiana hay una distancia tan grande como la que separa a la biología molecular de la ecología. Para los ecólogos sería una tarea imposible formular las regularidades de los sistemas que estudian empleando el vocabulario de los nucleótidos.

Aun cuando admitiéramos que la aparición de intencionalidad en un sistema de cómputo es bastante misteriosa, tendremos que admitir también que lo es aún más en un sistema bioquímico. Es más probable que sean las propiedades computacionales, y no las bioquímicas, las que finalmente se acerquen al secreto de la intencionalidad y la conciencia.

Hemos sido los psicólogos cognitivos, y no los bioquímicos, los que, al menos hemos podido describir, en un lenguaje de procesamiento, propiedades de ella tales como las de «proyectar esbozos de 2 1/2 D» (Jackendoff, 1987), «seleccionar información», «auto-programar» (Baars, 1988), etc.

Como señala Margaret Boden (1990), la intencionalidad y la conciencia son conceptos complejos y problemáticos en sí mismos, y no pueden compararse con los que usa Searle como marcos de referencia: cuando hablamos de la fotosíntesis tenemos muy claros sus productos, tales como los hidratos de carbono, y sabemos distinguirlos bien de otros, tales como las proteínas, pero «ninguna teoría de la intencionalidad se acepta sin problemas, a diferencia de lo que sucede con la química de los carbohidratos» (p. 93). La «síntesis química» de la intencionalidad es mucho más oscura que la de las proteínas: «tenemos buenas razones para suponer que las neuroproteínas sirven de soporte a la intencionalidad, pero no tenemos la menor idea acerca de *cómo* —*en tanto que* neuroproteínas— son capaces de hacerlo» (Boden, 1990, p. 93). Si algo sabemos, es que son las propiedades del sistema nervioso, como procesador de información (propiedades tales como la facilitación, la inhibición, la transmisión de mensajes) las que tienen algo que ver con el asunto.

La negativa de Searle a aceptar la invitación de Turing a deshacer, en un cierto nivel de abstracción, los límites conceptuales mentes (Fm)/máquinas (-Fm), parece estar fundamentada en un desdibujamiento previo de los límites conceptuales entre una bioquímica (¿por qué no llevar más allá la reducción, llegando por ejemplo a la física de partículas elementales?) de la que desaparece el concepto, quizá bioquímicamente útil, de información, y una psicología intencional que se sitúa muy lejos de aquella.

Como el mismo Searle dice, se trata de hacer compatible un «fisicalismo ingenuo» con un «mentalismo ingenuo». Si se quita el primer término, todo queda en una tesis eliminacionista (lo mental admite una reducción lingüística completa al vocabulario de la neurofisiología), bastante semejante a la de Churchland (1986). Las críticas de Searle a la ciencia cognitiva del paradigma clásico (propensión dualista, formalismo excesivo, inconsistencia del concepto de símbolo, descuido de problemas relacionados con el significado intencional, etc.) son muy dignas de tenerse en cuenta. Su solución positiva no parece serlo tanto: se levanta sobre una confusión considerable de planos de explicación autónomos.

En suma: ni es correcto predicar de las sinapsis que tienen intencionalidad, ni sensato decir que «Basilio votó a los conservadores a causa de una condición de su hipotálamo» (Searle, 1985, p. 49), ni parece necesario abandonar, por ahora, la pretensión de los psicólogos cognitivos de hacer una ciencia objetiva de la mente. Esa ciencia ha producido ya conocimientos empíricos importantes, que frecuentemente no resultan intuitivos para el sentido común. La reconstrucción de los mecanismos computacionales de la visión, por ejemplo (vid. Marr, 1982), permite conocer un plano de leyes y relaciones que son tan invisibles para el ojo de la mente como los cuantos de luz, como tales, lo son para los del cuerpo. Tales leyes son propiamente cognitivas y no pueden reducirse, sin pérdida de significado, a las vinculadas con la neurofisiología del quiasma óptico o las propiedades bioquímicas de la redosopina.

Por todo ello, el pronóstico negativo de Searle sobre el futuro de la psicología cognitiva parece fundamentarse en un diagnóstico poco preciso. Podríamos hacerle, así, esa observación tan expresiva de uno de nuestros clásicos: «Los muertos que vos matáis, gozan de buena salud».

4.4. La alternativa literal del Paradigma C-R

La segunda alternativa al desafío de Turing ha sido la desarrollada principalmente por investigadores del *«Massachusets Institute of Technology», como Chomsky (1980) y Fodor (1975), que pueden considerarse creadores de la ciencia cognitiva. También han contribuido mucho a su fundamentación otros investigadores, como Newell y Simón (1976) y Zenon Pylyshyn (1984, 1989), que se sitúan en la imprecisa franja que está entre la Inteligencia Artificial y la psicología cognitiva propiamente dicha. Se basa en un enfoque funcionalista, desarrollado por filósofos como Hillary Putnam (1960), Fred Dretske (1988) y el propio Jerry Fodor (1988).

Todos estos investigadores mantienen la idea de que puede definirse un plano cognitivo autónomo, en el cual la ruptura de la barrera tradicional «mentes (Fm)/ máquinas (-Fm)» admite una interpretación *literal* y no sólo metafórica. Por consiguiente, creen también que es posible y deseable el desarrollo de *una* sola ciencia cognitiva unitaria, en la que se estudian las propiedades comunes que poseen todos los sistemas capaces de conocer, con independencia de que estén fabricados de proteínas o de tubos de silicio. Esa ciencia cognitiva sería, a la vez, consecuencia y fundamento de las ciencias particulares que tratan sobre temas

relacionados con el conocimiento: en especial, partes de la lógica y la lingüística, la psicología cognitiva y la I.A. El objetivo de la ciencia cognitiva integrada sería el de *abstraer* los principios generales a esas otras ciencias, mediante un doble proceso de reflexión rigurosa e inducción de lo que hay de fundamental y común en los resultados de tales ciencias.

Lo más general y común que puede decirse de objetos tales como los cálculos lógicos, la definición de las estructuras del lenguaje, los programas de ordenadores y los procesos cognitivos de las mentes es que todos ellos consisten en sistemas de reglas tales que permiten realizar cómputos sobre representaciones simbólicas (Chomsky, 1980). En ese sentido, el juego de Turing puede aceptarse literalmente: todas las mentes y algunas máquinas pueden considerarse, en efecto, como elementos de una sola clase común, la de las «cosas cognoscentes» o dispositivos que consumen, organizan y tratan conocimiento:

De la misma forma que el dominio de la biología incluye algo así como «todas las cosas vivientes» (algo cuya estricta definición probablemente sea imposible fuera de la ciencia biológica) —dice Zenon Pylyshyn—, es posible que el dominio de la ciencia cognitiva sea algo del estilo de «las cosas cognoscentes» o, tal y como Miller (1984) ha expresado vividamente, los «informávoros» (1984, ed. esp. 1988, p. 10).

Por la misma razón por la que los avances biológicos contribuyen a un mejor conocimiento del hombre como ser vivo, los de la ciencia cognitiva integrada ayudan a su comprensión como ser cognoscente. La categoría de los informávoros estaría formada, como mínimo, por los vertebrados superiores y algunos sistemas informáticos. Es explicable que a las personas las moleste un poco, al principio, verse incluidas en la misma categoría que algunas máquinas. Eso ya sucedió cuando la ciencia biológica las incluyó en la misma categoría que los otros primates. Pero la ciencia tiene que avanzar en oposición a ese tipo de resistencias antropocéntricas, de las que su historia está llena. Zenon Pylyshyn (1988) hace una observación muy pertinente para ayudarnos a pasar el trago:

Para que la perspectiva de ser primos lejanos de los ordenadores no nos parezca tan extemporánea como lo fue en su momento la que nos presentaba como descendientes de los grandes monos, debemos recordar en todo momento que sólo se trata de formas de clasificación de los individuos que se adoptan con el

propósito de descubrir algunos de sus principios operativos. Después de todo, para averiguar cómo nos movemos en respuesta a las fuerzas físicas, se nos clasifica junto con las rocas, los átomos y las galaxias. Ninguna clasificación —incluyendo las que hablan de «descendientes», «seres sentientes» o «ángeles caídos» —puede captar *todo* lo que es exclusivamente humano /.../. Pero cada una de las aproximaciones nos da una visión especial de algún aspecto de nuestra naturaleza (ibid., p. 11).

La fusión «mente-máquina simbólica» en una sola categoría Fm se toma, por tanto, *en serio.* No es sólo una metáfora que deba ser abandonada después de que haya cumplido funciones de provocar el pensamiento y favorecer la investigación empírica, sino un principio esencial de la ciencia cognitiva unitaria, que sirve como fuente constante de hipótesis heurísticas y del que deben sacarse las consecuencias racionales que se deriven de él, sean éstas las que sean. En este sentido, hay que reconocer un punto importante a favor de los defensores del paradigma C-R: que han sido ellos, Jerry Fodor, Noam Chomsky, Zenon Pylyshyn, Hillary Putnam, etc, los que han articulado, en los últimos años, un pensamiento más riguroso en torno a los procesos de conocimiento. Ese rigor conceptual, unido a la evidencia manifiesta de los avances en las tecnologías del conocimiento, proporciona una gran fuerza, y coherencia, a sus argumentos y permite hablar de la existencia de un paradigma, unitario e interdisciplinar al mismo tiempo, de explicación del conocimiento: es a este enfoque al que hemos llamado —y seguiremos haciéndolo— «paradigma C-R» (de los cómputos sobre representaciones) a lo largo de estas páginas.

En numerosos escritos de los teóricos del paradigma C-R, se pone de manifiesto esa decisión firme de atenerse a todas las consecuencias que pueda tener la comprensión literal de las mentes y las máquinas abstractas como miembros de una misma clase común de sistemas computacionales que actúan sobre un medio de representaciones simbólicas. Así, en la presentación de *Computación y conocimiento,* Pylyshyn comenta:

Fodor y Newell, que practican estilos intelectuales casi completamente opuestos, comparten, sin embargo, algo que no cabe duda que ha afectado a la redacción de este libro: se encuentran entre los pocos estudiosos de la ciencia cognitiva que se conducen de acuerdo con el principio de que si creemos que P, y si creemos que de P se sigue Q, entonces, si P no parece algo descabellado, tenemos la obligación intelectual de tomar en serio la posibilidad de que Q sea

verdadero. En esto consiste tomar las ideas en serio, y esto es lo que yo defiendo que tenemos que hacer con la idea del conocimiento como computación (ed. española de 1988, p. 22).

Veamos algunos ejemplos: la «P» que comparten, por definición, los defensores del paradigma C-R es la que Newell y Simón (1976) llaman «Ley de estructura cualitativa» fundamental de la psicología cognitiva (y que jugaría en ésta un papel muy semejante al de las leyes básicas de estructura cualitativa en otras ciencias, como la doctrina celular en Biología, la tectónica de placas en Geología, etc), a saber, que la mente capaz de conocimiento y acción inteligente es un sistema con las siguientes propiedades:

(1) Computa símbolos capaces de designar objetos y de ser interpretados en forma de procesos.

(2) Como tal sistema que computa símbolos, la mente está determinada en su funcionamiento por sus representaciones. Es, como decía Descartes, «un sistema que obra por conocimiento y no por la disposición de sus órganos».

(3) En tanto que estas representaciones son, en último término, estados físicos el sistema, el funcionamiento mental, en el nivel molecular del «lenguaje máquina», se atiene a leyes causales físicas.

(4) Puesto que ese funcionamiento es computacional, la mente se guía, de forma determinista, por la naturaleza sintáctica y formal de los símbolos.

Esta Ley fundamental de estructura cualitativa da lugar a problemas conceptuales —ya nos hemos detenido en algunos— para explicar ciertos fenómenos aparentes de representación, conciencia, aprendizaje, desarrollo e influencia de los contenidos en los procesos cognitivos. Pero, si se toman en serio las premisas 1 a 4, habrá que arrostrar las consecuencias Q, y cuestionar el significado de los fenómenos o, mejor, proponer teorías completamente consistentes con las premisas anteriores. Veamos algunos ejemplos:

(1) En relación con los fenómenos de *representación,* no resulta fácil comprender la funcionalidad de las imágenes mentales desde el paradigma C-R, dado que éstas son continuas, «densas» y analógicas, mientras que un sistema *simbólico* de representación es discreto, analítico y compuesto de símbolos atómicos articulados en combinaciones por medio de una sintaxis. En suma, las imágenes no son, en sí mismas, un *medio* de computación. Y, si no lo son, su valor *funcional* como representaciones

es muy dudoso. En consecuencia, ha habido una larga e influyente polémica, en psicología cognitiva, sobre el valor funcional de las imágenes. Frente a las pretensiones de investigadores de propensión más empirista, como Cooper y Shepard (1973), Stephen Kosslin (1980), Paivio (1971), etc, que han pretendido demostrar la funcionalidad y autonomía estructural de las representaciones por imágenes, Pylyshyn (1973, 1978, 1980, 1981, 1984) ha mantenido tenazmente la tesis de que no son *inteligibles,* desde un modelo C-R, los supuestos de autonomía y funcionalidad, proponiendo la necesidad de reducir las explicaciones sobre imagen a explicaciones proposicionales (i.e. simbólicas) del conocimiento.

(2) Con respecto a la *conciencia,* Ray Jackendoff (1987), por ejemplo, realiza un análisis de sus relaciones con la mente computacional, en el más impecable estilo del paradigma C-R. Formula la que denomina «Teoría del nivel intermedio», según la cual hay diversas modalidades (por no decir «módulos») de conciencia que corresponden a proyecciones de estructuras producidas en fases medias de procesamiento de los que serían procesos «completos» en las modalidades correspondientes (por ejemplo, la «conciencia visual» correspondería a proyecciones del esbozo 2 1/2 d (representaciones en «dos dimensiones y media»), que se produce antes de la imagen de 3 dimensiones en los procesos visuales, la lingüística sería una imagen intermedia de la estructura fonológica real, etc). Estas proyecciones serían el resultado de procesos de selección de materiales de la memoria a corto plazo, enriquecidos por la atención. Sin embargo, Jackendoff (1987) sigue defendiendo la tesis C-R de la «ineficacia causal de la conciencia», y considera, por ejemplo, que «la introspección es un proceso puramente computacional, en general poco diferente de otros procesos de translación», de tal manera que «el problema mente-mente sigue aún sin resolverse, aunque al menos no queda peor de lo que estaba» (p. 317).

(3) En cuanto al aprendizaje, es posible admitir y explicar, desde una perspectiva C-R, las que podríamos denominar «formas superficiales» de aprendizaje (tales como las que implican relacionar unos conceptos con otros, refinar y modificar esquemas previos, etc). Sin embargo, es muy difícil dar cuenta de «formas profundas»: ¿cómo puede alcanzar la sintaxis subyacente (la «sintaxis última») de un sistema simbólico mayor poder de cómputo que el que ya posee de antemano? La posición innatista de Chomsky (1985) y Fodor (1975) no es un simple prurito de respeto a la tradición racionalista que defienden, sino una exigencia profunda del enfoque C-R.

Los sistemas propiamente *simbólicos* que manejan información de manera exclusivamente lingüística (i.e. ignorando las posibles relaciones entre los símbolos y procesos de naturaleza continua) tienen muy limitadas sus posibilidades endógenas de auto-complicación (Carello, Turvey, Kugler y Shaw, 1982, Pattee, 1977, 1982. 1984; Riviére, 1985, 1986). Por eso Fodor (1983), que ve con mucha lucidez este problema, formula la tesis provocativa de que *el aprendizaje conceptual es imposible,* y señala, en la introducción castellana a *El lenguaje del pensamiento* (1975, cit. ed. esp., 1984) lo siguiente:

Algunas de las doctrinas predilectas de LDP han adquirido cierta notoriedad —en especial el nativismo comprensivo adoptado en el capítulo 2—, pero, de hecho, no son más que la consecuencia lógica de una aceptación escrupulosa de la explicación computacional/representacional de la mente cognitiva. Si, por ejemplo, el aprendizaje es un proceso computacional —un proceso definido sobre símbolos mentales—, estos símbolos mentales tienen que proceder de alguna parte; si son aprendidos, los símbolos que median *ese* aprendizaje deben proceder, a su vez, de algún lugar. Tarde o temprano, ese retroceso deberá detenerse. Cualquiera que sea el punto en que nos detengamos, se acepta ipso facto que *no* es aprendido. La idea de LDP era que, dado que el nativismo comprensivo forma parte del coste que hay que pagar para tener una teoría computacional/representacional de la mente, no queda otro remedio que pagar tal precio (p. 20).

(4) Pocas líneas después, Fodor reconoce también abiertamente las dificultades del paradigma C-R para explicar la influencia y significación de los aspectos relacionados con los *contenidos* y la *semántica* de las representaciones mentales, aunque él mismo da después algunos pasos importantes, en *Psychosemantics* (1988) para tratar de solucionar este problema.

Creo que los comentarios anteriores bastan para ilustrar las características principales de esta segunda alternativa al reto de Turing, a la que hemos llamado «Paradigma C-R». Es la que corresponde a lo que otros investigadores (vid. De Vega, 1982) han denominado «versión fuerte» de la metáfora del ordenador. En general, ha sido defendida por filósofos de la mente, lingüistas, lógicos y especialistas en I.A. más que por psicólogos dedicados a la investigación empírica (aunque también hay, entre éstos, defensores de esta versión fuerte). En un estilo muy coherente con el de la tradición racionalista, que tanto ha influido en este paradigma, sus proponentes suelen tender a preferir las conclusiones que son deductivamente coherentes con sus propias premisas a aquellas otras

que parecen más compatibles con supuestas evidencias fenoménicas o datos empíricos. En muchos casos, prefieren un buen razonamiento, una buena simulación, o incluso un buen programa de I.A. (aunque no pretenda ser simulación de la mente) a un cúmulo de datos experimentales. Estas últimas son tendencias estilísticas del pensamiento en el paradigma C-R, más que principios fundamentales que definan a ese paradigma.

La aportación del paradigma C-R a la psicología y la ciencia cognitiva ha sido, en todo caso, de una enorme importancia. Implica una exigencia de rigor conceptual y precisión algorítmica en los modelos de la mente muy beneficiosa para la psicología. Sirve, una y otra vez, de semillero de problemas profundos que debe plantearse cualquier modelo psicológico del conocimiento. Facilita el desarrollo de modelos de la mente que aúnan una amplia generalidad explicativa y una minuciosa precisión computacional, lo cual posibilita la realización de simulaciones (por ejemplo, la teoría ACT de Anderson, 1976, 1983). Sin el paradigma C-R la psicología cognitiva hubiera sido decisivamente diferente y, muy probablemente, mucho menos rigurosa e interesante. Algunas obras de los investigadores situados en el marco de este paradigma se encuentran, probablemente, entre las creaciones intelectuales más significativas de nuestra época. En todos estos sentidos, ha merecido la pena que se tomara en serio el juego de Turing.

Sin embargo, la psicología es una ciencia experimental y que, con frecuencia, tiene que hallar soluciones de compromiso entre la coherencia conceptual y la consistencia empírica. Las limitaciones impuestas por una interpretación literal del reto de Turing pueden ser excesivas para muchos psicólogos, que prefieren el sentido metafórico y débil de la disolución de la frontera entre las mentes y las máquinas simbólicas.

4.5. El juego de Turing como metáfora: la psicología del procesamiento de la información

Como señala De Vega (1982), «el mero reconocimiento de una analogía funcional entre dos sistemas está cargado de ambigüedad. Prueba de ello es que la metáfora computacional ha dado lugar a vías de investigación sumamente dispares. Se puede establecer, *grosso modo,* una versión débil y una versión fuerte de la analogía/.../. La versión fuerte y la versión débil de la metáfora computacional constituyen prácticamente dos subparadigmas independientes, pese a su inspiración común en

dicha analogía» (pp. 64-65). La versión débil corresponde a la interpretación del juego de Turing realizada por la mayor parte de los psicólogos que se sitúan en el marco de los llamados —muy inadecuadamente— «modelos de procesamiento de la información».

La posición de la mayoría de los psicólogos que construyen teorías de la mente con el enfoque del procesamiento de la información implica un compromiso mucho menor que en el paradigma C-R, con respecto a la propuesta de disolución de los límites mente/máquina, para la atribución de predicados Fm. Algunos de los proponentes del paradigma C-R han señalado la diferencia con claridad. Pylyshyn (1980), por ejemplo, comenta lo siguiente:

a pesar del empleo generalizado de una terminología computacional (por ej., términos como «almacenamiento», «proceso», «operación»), en gran parte de estos usos, los términos se emplean dándoles, como mucho, un contenido metafórico. Ha habido una resistencia a tomar la computación como una descripción *literal* de la actividad mental, y no sólo como una metáfora con un valor meramente heurístico. Desde mi punto de vista, la falta de consideración de la tesis de la computación en un sentido literal ha posibilitado una amplia gama de actividades realizadas bajo la rúbrica de la «teoría del procesamiento de la información», algunas de las cuales representan una desviación significativa de las ¡deas que considero básicas en toda teoría computacional de la mente (p. 114).

¿Cómo puede caracterizarse esta interpretación débil de la propuesta de Turing, en oposición a la más literal? A mi parecer, tendría los siguientes rasgos:

(1) Adopta el *vocabulario teórico* de la ciencia de la computación para referirse a las actividades cognitivas de la mente: ésta realiza cómputos, accede a informaciones almacenadas en la memoria, está más o menos limitada en su velocidad de procesamiento, es —o no— un procesador serial, etc. La justificación última de ese vocabulario teórico reside en el intento de explicar determinados conjuntos de datos de conducta. La formulación de modelos que permitan simular en ordenador las actividades cognitivas estudiadas tiene un valor complementario importante en algunos casos, pero los criterios últimos de justificación de los modelos teóricos del Procesamiento de la Información suelen ser los fenómenos empíricos que se observan en el estudio experimental de la conducta de las personas, cuando éstas realizan actividades cognitivas.

(2) El estilo de pensamiento con el que se formulan los modelos P.I. de la metáfora débil es, frecuentemente, *inductivo* más que *deductivo*.

El investigador se guía muchas veces por un «marco conceptual» muy laxo, que contiene un vocabulario de procesamiento; sus hipótesis suelen basarse en resultados de observaciones empíricas anteriores y, con menor frecuencia, en deducciones rigurosas de la interpretación computacional de la mente. Los contrastes de tales hipótesis consisten en datos obtenidos en situaciones experimentales que respetan una larga tradición de desarrollo de instrumentos metodológicos de justificación (diseños experimentales, procedimientos de análisis estadístico, etc) en una ciencia empírica como es la psicología. La construcción de nuevos modelos, y la sustitución, modificación o refinamiento de los anteriores se fundamenta en la base proporcionada por: (a) una amplia gama de conocimientos anteriores, sobre el campo estudiado, obtenidos del estudio empírico de la conducta, (b) los datos propios (generalmente, latencias o errores de los sujetos) y, (c) más secundariamente, por intuiciones que provienen del propio vocabulario computacional.

(3) Como señalan Johnson-Laird (1975) y De Vega (1982), la psicología del procesamiento de la información y el paradigma C-R tienden, en cierto modo, a ajustarse a criterios de justificación de la verdad científica que son sutilmente diferentes: la primera emplea, sobre todo, el criterio de la correspondencia entre las predicciones que se derivan de los modelos teóricos y las observaciones de conducta obtenidas en situaciones controladas. El criterio del paradigma C-R es, más bien, la coherencia entre los enunciados teóricos. Así, éste acentúa la función sintáctica de las teorías (lo que es muy coherente, en un modelo de computaciones guiadas por la sintaxis de los símbolos sobre los que se realizan), mientras que la psicología de P.I. acentúa su valor semántico (De Vega, 1982).

(4) Muy en relación con el punto anterior, sucede que, en la medida en que disminuyen las exigencias de coherencia formal de los modelos P.I., en comparación con los del paradigma C-R, aumentan sus exigencias *semánticas:* se admiten mejor algunos «agujeros explicativos» en los modelos que las faltas de cuidado, rigor y justificación de las tareas y datos en que tales modelos se fundamentan. Con independencia de que las teorías alcancen el mayor grado de coherencia conceptual posible, tienen que estar basadas en observaciones y datos. Del mismo modo que Pylyshyn (1980) ve con alguna desconfianza la falta de coherencia «sintáctica» de muchas de las teorías de los psicólogos del Procesamiento de la Información, éstos tienden a considerar con alguna suspicacia la «base semántica» de observaciones que emplean frecuentemente los teóricos

del paradigma C-R: intuiciones lingüísticas (Chomsky, 1965), protocolos de «pensamiento en voz alta» (Newell y Simon, 1972, Ericcson y Simón, 1984; vid. revisión en Simón y Kaplan, 1989) o intuiciones generales sobre la forma de procesos mentales (Schank y Abelson, 1977).

(5) La interpretación débil de la metáfora computacional se basa, generalmente, en el supuesto de una correspondencia vaga entre componentes fijos del ordenador y la mente: un procesador central, que regula el flujo de información, y cuenta con una memoria de trabajo; una memoria más permanente (o «a largo plazo») y diversos periféricos de entrada (sistemas sensoriales, en el caso de la mente) y salida (sistemas motores). Con frecuencia, se emplean términos del vocabulario teórico de la computación para mantener tesis, acerca del funcionamiento de cada uno de esos subsistemas y el sistema conjunto, que no son fáciles de compatibilizar con los supuestos actuales de la teoría de la computación.

(6) En la mayor parte de los casos, los psicólogos del P.I. no llegan a construir modelos algorítmicos completos de los procesos que estudian. Muchas de sus investigaciones permiten sólo decir si ciertas variables influyen o no influyen en tales o cuales procesos supuestos, al observar cómo condicionan la conducta. Pero no permiten establecer, paso a paso, las transformaciones sobre representaciones que realizan procesos bien definidos de cómputo. En estas condiciones, el empleo de la metáfora computacional sólo puede ser débil. La razón es muy obvia: el concepto de «cómputo» sólo tiene sentido en el marco de un algoritmo completo, de un «procedimiento efectivo» que efectivamente realiza de *forma automática* una determinada tarea. No existen cómputos cuando no existen procedimientos efectivos. Y no existen procedimientos efectivos cuando no se definen algoritmos completos. Por último, los algoritmos no son completos cuando no especifican todos y cada uno de los pasos que se dan en un sistema para generar una conducta.

(7) La consecuencia de lo anterior es muy clara: en la mayor parte de los casos, los psicólogos del Procesamiento de la Información realizamos una operación algo sospechosa. Esta consiste en que nos beneficiamos de las fuertes connotaciones de rigor, cientificidad y eficacia de los términos del vocabulario de la computación, para formular y defender modelos que no son, en ningún sentido rigurosos, *modelos computacionales,* desde el momento en que no son completos, ni especifican procedimientos efectivos para realizar las tareas que vemos realizar a los sujetos cuyas conductas estudiamos.

(8) Finalmente, la conjunción del estilo inductivo de construcción teórica, el rigor «minimista» exigido a los datos, la dificultad de formular modelos algorítmicos completos de la mente, los fuertes residuos operacionalistas de la psicología empírica —con lo que éstos suelen implicar de prevención ante las formulaciones teóricas que se alejen más de dos pasos de los datos empíricos—, y el escaso «valor ecológico» de muchas de las tareas estudiadas por los psicólogos del Procesamiento de la Información, da lugar a una psicología más bien *mosaica,* fragmentada en multitud de modelos también minimistas, capaces de explicar conjuntos limitados de observaciones sobre situaciones, muy estrechamente delimitadas, y difíciles de encajar en concepciones integradas de la mente.

Un punto diferenciador importante entre los psicólogos empíricos del Procesamiento de la Información y los teóricos del paradigma C-R es el hecho de que aquellos, al acentuar la importancia de la función semántica de correspondencia entre teorías y datos, *no* están dispuestos a asumir consecuencias que parecen deducirse de la interpretación literal de la cognición como computación, en los casos en que tales consecuencias no resultan compatibles —al menos en un primer análisis— con fenómenos empíricos comprobados. Los ejemplos de esta actitud podrían multiplicarse, pero nos referiremos sólo a dos temas: la conciencia y las imágenes mentales.

Baars (1988), en relación con el obstinado problema de la conciencia, realiza una amplia revisión de las aportaciones de la psicología del procesamiento de la información a la comprensión de su función. A diferencia de Jackendoff (1987), situado en el paradigma C-R, Baars no asume en absoluto la tesis de la ineficacia causal de la conciencia, que parece deducirse de una concepción de la mente como autómata simbólico:

> una premisa básica de este libro es —dice— que, como cualquier otra adaptación biológica, la conciencia es *funcional* /.../, quizá su función más fundamental es /.../ la capacidad de alcanzar un punto óptimo de equilibrio entre organización y flexibilidad. Las respuestas organizadas son muy flexibles en situaciones bien conocidas, pero cuando se dan situaciones nuevas, la flexibilidad es una ventaja. Desde luego, la arquitectura global del espacio de trabajo está diseñada de manera que las soluciones «prefabricadas» se hagan accesibles automáticamente en situaciones predicables, y también de forma que puedan combinarse muchas fuentes diferentes de conocimiento en las circunstancias impredictibles (Baars, 1988, pp. 347-348).

En su defensa de la tesis de la funcionalidad y eficacia causal de la conciencia, Baars (1988) emplea, en todo momento, un vocabulario computacional estricto, que podría producir la impresión engañosa de que se sitúa en el mismo paradigma que Jackendoff. El uso de un vocabulario teórico común tiende, en efecto, a desdibujar las importantes diferencias estilísticas que existen entre el paradigma C-R y la psicología del procesamiento de la información. Sin embargo, un análisis más detenido de las formas de hacer ciencia que se manifiestan en los textos de Jackendoff (1987) y Baars (1988) permite entrever las distinciones que se han apuntado en este apartado: el segundo parte de una base empírica más amplia, y de orígenes más diversos; tiende a defender las tesis que pueden inducirse desde esa base «semántica»; el primero selecciona los materiales empíricos üe entre los más coherentes con una visión computacional-representacional, acentúa la integridad sintáctica de su modelo, y fundamenta muchos de sus argumentos en deducciones que se derivan de la base teórica.

En el estudio psicológico de la representación del conocimiento se ha producido, con mayor claridad que en ningún otro terreno, la confrontación entre estos dos estilos representados por el paradigma C-R y la psicología empírica del procesamiento de la información. Investigadores como Paivio (1971), Kosslyn (1980), Shepard y Metzler (1971), Cooper y Shepard (1973), Richardson (1980), Huttenlocher (1968) y Finke (1989) han demostrado que las imágenes mentales:

(a) son representaciones funcionales en actividades de comprensión, razonamiento y memoria.

(b) Poseen propiedades analógicas, que se manifiestan en fenómenos tales como los efectos de ángulo de rotación y «distancia mental» en tareas simples de solución de problemas (cuando los sujetos tienen que reconocer si una figura es o no la representación rotada de otra, se produce una relación lineal entre el tiempo que tardan y el ángulo de rotación),

(c) Preservan parcialmente las relaciones percibidas entre los objetos.

(d) Son susceptibles de transformaciones mentales continuas y no discretas (rotaciones, expansiones, etc).

(d) Consumen recursos comunes con la percepción (De Vega, 1988).

Como veíamos en el apartado anterior, la funcionalidad aparente de las imágenes mentales, y las propiedades estructurales que éstas tienen, así como la casi inevitable relación del significado del concepto de imagen con la fenomenología de la conciencia, resultan muy difíciles de explicar desde los supuestos del paradigma C-R: las imágenes no se corresponden con los medios lingüísticos de computación de los sistemas conocidos, no se organizan en símbolos atómicos arbitrarios con una sintaxis definida por reglas discretas, y son sensibles a los significados «proposicionales» subyacentes a su formación y transformación.

La polémica consecuente a estas consideraciones, abierta por Pylyshyn (1973), es un buen ejemplo tanto de las posiciones características de los defensores del paradigma C-R y los psicólogos de propensión más empírica, como del gran valor que ha tenido para la psicología cognitiva esta dialéctica.

Las críticas de investigadores como Pylyshyn (1973, 1978, 1980, 1981, 1984) y Fodor (1975) han constituido un acicate constante para el diseño de experimentos cada vez más ingeniosos encaminados a demostrar tanto la funcionalidad como las propiedades estructurales de las imágenes (vid. revisiones de Kosslyn, 1980, Rivière, 1986, y Finke, 1989). Además han favorecido el desarrollo de modelos muy precisos de simulación de fenómenos de imagen en sistemas de cómputo (Kosslyn y Schwartz, 1977) y han obligado a refinar los conceptos teóricos sobre las imágenes, al intentar salvarlos de las críticas agudas y rigurosamente fundamentadas de sus detractores (Kosslyn y Pomerantz, 1977; Paivio, 1977). De este modo, la aparente inconsistencia entre los fenómenos observables relacionados con las imágenes y las consecuencias deducibles de un modelo computacional-representacional estricto ha jugado un papel decisivo en el desarrollo de la teoría y la investigación empírica sobre la representación (del mismo modo que, en el desarrollo del niño, tiene una gran importancia la «toma de conciencia» de las contradicciones entre sus esquemas conceptuales y los fenómenos empíricos; vid. Piaget, 1974).

El ejemplo del estudio de las imágenes mentales constituye una ilustración clara de la utilidad de fondo que ha tenido para la psicología cognitiva el «choque de subparadigmas», la oposición entre la versión más literal y la más metafórica de la nueva categoría «mentes/máquinas». El estilo teórico de los defensores del paradigma C-R ha servido como acicate para el desarrollo del conocimiento empírico y de formulaciones conceptuales rigurosas sobre los fenómenos de imagen. La psicología tiene un compromiso empírico ineludible, y así debe ser. Sin embargo, ha tendido a ser durante muchos años (y salvando excepciones como la del clásico citado al final del último párrafo) una ciencia demasiado pa-

cata en el uso de la imaginación teórica, y propensa a aceptar dosis altas de inconsistencia en los modelos de la mente. Ello se ha relacionado con su estilo empirista de hacer ciencia, por simples procesos de acumulación de observaciones e inducción consiguiente.

Es cierto que los psicólogos tenemos que tener muy presentes las *diferencias sustantivas* que existen entre mentes y ordenadores: aquellas se enfrentan a la difusa y rica diversidad de los ambientes naturales a los que se adaptan, y éstos no; emplean conceptos difusos y de estructura probabilística, se guían por motivaciones y se colorean por emociones que no tienen los ordenadores; son resultado de las relaciones entre las conductas y el medio (por lo que tienen un componente de «implicación eferente» aún poco estudiado y mal comprendido por los psicólogos cognitivos), son capaces de auto-programación espontánea y aprendizaje (De Vega, 1982), son dispositivos *sociales* y no sistemas solipsistas (Seoane, 1984), tienen naturaleza intencional (Searle, 1983) y son conscientes de buena parte de los objetos con los que tratan. Todo eso es cierto y debe ser tenido en cuenta.

Sin embargo, sería deseable que los psicólogos fuéramos algo más propensos a analizar profundamente los «postulados terminales» en que se basan nuestros términos teóricos, objetos de investigación y conceptos subyacentes de la mente. A mi parecer, resulta criticable la tendencia, demasiado frecuente, a hacer uso de las ventajas connotativas del vocabulario de procesamiento, sin analizar ni asumir a fondo sus posibles consecuencias racionales. ¿Son éstas tan graves como para abandonar el vocabulario que la psicología del conocimiento ha empleado más en los últimos treinta años? Por ahora, no lo parece.

El enfrentamiento entre las consideraciones de origen empírico de los psicólogos del Procesamiento de la Información y las razones aducidas por los defensores más lúcidos del paradigma C-R ha sido, en términos globales, muy positivo para el desarrollo de la psicología cognitiva vagamente basada en los modelos actuales de computadores con un procesamiento serial de cadenas simbólicas, a la manera del modelo de von Neumann. El vocabulario generado desde ese modelo ha tenido un valor histórico enorme para el desarrollo de una ciencia objetiva de la mente. Hasta hace muy pocos años, sucedía, además, que no existía un vocabulario alternativo que tuviera un valor heurístico y una precisión conceptual comparables. Sin embargo, en los últimos años, se ha desarrollado de forma efectiva un nuevo modo de hablar de la mente, al que debemos referirnos ahora.

4.6. Cambiar las reglas del juego de Turing: el enfoque conexionista

Las limitaciones de los modelos que conciben la mente como un procesador serial se han ido haciendo cada vez más evidentes a lo largo de los años setenta y ochenta. Aparte de las anomalías de la conciencia y el aprendizaje, a que ya nos hemos referido, y de la dificultad lógica para dar sentido al supuesto de los «símbolos inconscientes», sucede que esa mente serial presenta problemas importantes para enfrentarse a condiciones que son muy naturales en el medio adaptativo del hombre. Por ejemplo, las personas nos enfrentamos muy frecuentemente a situaciones en que es necesario tener en cuenta, a la vez, diversas restricciones simultáneas. Ello sucede cada vez que tenemos que hacer algo tan simple como tomar con los dedos una moneda que se ha «colado» detrás de un mueble, o leer un párrafo de un libro al que se le ha deteriorado parte de una página por haberse «pegado» a otra.

Además, ocurre con mucha frecuencia que nos vemos en la situación de reconstruir o recordar una información a partir de una parte de ella que incluso puede no ser del todo correcta. El medio sobre el que se realizan nuestros procesos cognitivos no tiene la limpia y rigurosa estructura de los símbolos que se le proporcionan, a través de un teclado, a un ordenador. Es más diverso en sus demandas, más impreciso en sus contornos, más ruidoso en muchos de sus contenidos informativos —oscurecidos por procesos de degradación, solapamiento, etc—, más dinámico que el del ordenador.

Nuestro sistema cognitivo soporta, con gran entereza, condiciones de degradación parcial de la información con la que trata, que pueden provenir tanto del medio externo como de su propio medio interno (cambios de atención, estados de somnolencia, variaciones en el riego cerebral, etc). Es, por otra parte, especialmente hábil para captar, con rapidez, significados globales de las situaciones. Sus esquemas y conceptos poseen una gran flexibilidad, a diferencia de los conceptos y esquemas que se representan, en forma de listas, redes, etc, en las memorias de los ordenadores digitales de la familia «von Neumann» (ordenadores seriales, que funcionan de manera digital, a partir de órdenes que se especifican en lenguajes simbólicos de alto nivel). Además, nuestro sistema cognitivo emplea estrategias heurísticas que son sutilmente diferentes de las más algorítmicas y exhaustivas que usan esas máquinas simbólicas tan poderosas.

En suma, cualquier modelo cognitivo realista de la mente tiene que

enfrentarse —como mínimo— a varias exigencias difíciles: (1) ser, a la vez, *estructurado* y *flexible,* (2) afrontar condiciones de degradación parcial del medio interno (organismo) y externo, (3) dar cara a demandas múltiples y simultáneas, (4) asignar significados rápidos, sin necesidad de recurrir a algoritmos exhaustivos, (5) completar patrones conceptuales, perceptivos, etc, sin contar con toda la información *lógicamente* necesaria para hacerlo, (6) acceder a sus conocimientos por vías diversas, y tener sistemas de memoria en que éstos puedan recuperarse en virtud de sus contenidos, y (7) modificarse adaptativamente en función de las experiencias previas.

Estas consideraciones, propiamente psicológicas, han sido las que han llevado, en los últimos años, a algunos psicólogos —entre los que se cuentan D. Rumelhart, G. Hinton, J. McClelland, J. Anderson, J. Feldman, D. Ballard, P. Smolensky, T. Sejnovski y D. Zipser— a recuperar un proyecto que se había quedado un tanto olvidado en un pliegue de la historia, y del que hablábamos unas páginas atrás: el de formular teorías cognitivas que, al mismo tiempo, sean respetuosas con esas propiedades y exigencias psicológicas, a lás que acabamos de referirnos, y conformes, en un sentido abstracto, con aspectos esenciales del funcionamiento y la estructura del sistema nervioso. La historia ha venido a demostrar que las redes neurales de McCulloch y Pitts, Hebb, Rosenblatt y Minsky sufrieron, como las almas de las bestias y otras entelequias de Leibniz, «un largo aturdimiento» y no «una muerte en sentido riguroso». De este modo, han podido resucitar en los últimos años.

La condición esencial para la resurrección de la tradición conexionista ha sido trabajar sobre redes que, a diferencia de los perceptrones de Rosenblatt, no constaran sólo de una capa de unidades de in-put y otra de unidades de out-put, sino también de capas de unidades «ocultas» («hidden»), situadas entre las primeras y las segundas. En los modelos neo-conexionistas, esas unidades (que pueden concebirse como una especie de «neuronas abstractas») se definen por estados de activación y umbrales de out-put. Están conectadas mediante asociaciones unidireccionales que definen un «patrón de conectividad». Las conexiones entre las unidades pueden ser excitatorias o inhibitorias, y tienen un valor asignado, que corresponde a la fuerza de la conexión. El sistema formado por las unidades y sus conexiones funciona de acuerdo con ciertas reglas (en las que no debemos extendernos ahora) de propagación, activación y aprendizaje (estas últimas son elaboraciones a partir de la regla de Hebb a la que ya hicimos referencia: cuando dos unidades de una red neural están excitadas simultáneamente se incrementa la fuerza de conexión entre ellas).

¿Cómo funciona un sistema de este tipo?; ¿hasta qué punto es capaz de imitar, en aspectos relevantes, el funcionamiento de la mente?; ¿pasa el test de Turing? Rumelhart, McClelland y el grupo de investigación PDP (1986) han revisado un gran número de datos que demuestran que sí. Los modelos de procesamiento distribuido y paralelo, además de ser neurofisiológicamente más realistas que los clásicos, compuestos de secuencias de «computaciones-representaciones simbólicas», son capaces de pasar el test de pensamiento de Turing. Poseen esas propiedades «psicológicas» importantes que son tan difíciles de explicar desde la perspectiva clásica: son flexibles, hábiles para reconocer estímulos degrada^ dos, resistentes al deterioro de partes del propio sistema, capaces de acceder a información por vías diversas y guiadas por los contenidos, muy heurísticos en su modo de funcionar, competentes para reconocer patrones sin contar con toda la información lógicamente necesaria y, finalmente, inteligentemente sensibles a la experiencia previa, que les permite modificar su «patrón de conectividad».

Es importante destacar que los sistemas de procesamiento distribuido y paralelo sobre una red de unidades interconectadas, tales como los que se han esbozado a grandes rasgos en los párrafos anteriores, son muy diferentes de los sistemas clásicos del paradigma C-R, en varios aspectos sustantivos (y que son, ciertamente, los que definen la esencia de estos últimos):

(1) Su procesamiento no es serial sino paralelo. No realizan una secuencia de cómputos colocados en una fila temporal a lo largo de un camino en que sólo cabe un proceso después de otro. En las redes conexionistas se efectúan, por el contrario, numerosos cómputos simultáneamente. Ello plantea delicadas cuestiones relacionadas con la sincronía y la acomodación de los ritmos temporales de los procesos simultáneos (Rumelhart, MacClelland, y Grupo de Investigación PDP, 1986), y con la relación entre éstos y los procesos aparentemente más seriales de pensamiento; sin embargo, el paralelismo de los cómputos parece un supuesto más realista que la serialidad en relación con el funcionamiento del sistema nervioso.

(2) Los cómputos no consisten en transformaciones de representaciones simbólicas, sino en cambios cuantitativos que tienen lugar en los estados de activación y los patrones de conectividad de las redes. No se puede decir, en rigor, que respondan a reglas de una sintaxis. Más bien

UNIVERSIDAD DE ANTIOQUIA
BIBLIOTECA CENTRAL

corresponden a reglas de activación, propagación y aprendizaje que son «inherentes» al sistema.

(3) Las representaciones no son simbólicas en ningún caso. Pueden concebirse como patrones asignados localmente a unidades/conexiones del sistema o, más consecuentemente, como distribuidas en conjuntos de unidades y conexiones (Hinton, McClelland y Rumelhart, 1986); de ningún modo pueden entenderse como *estructuras simbólicas* explícitas en el sistema. Es decir, mientras que las representaciones de los modelos clásicos son *semánticamente transparentes*, las de los conexionistas no lo son (vid. Clark, 1989, para esta distinción). Así, los modelos conexionistas se desvían de la «Ley de estructura cualitativa» que define sustantivamente al modelo clásico del paradigma C-R, a saber: que la mente es un sistema físico-simbólico (Newell y Simón, 1976).

(4) Por consiguiente, los rasgos, los conceptos, las categorías, las proposiciones, los esquemas, los guiones, etc (las estructuras de conocimiento en una palabra) y las reglas, procesos, producciones, procedimientos, etc (i.e. las actividades sobre el conocimiento), no se representan explícitamente en los sistemas conexionistas. Estos no consisten en «programas escritos en lenguaje de alto nivel», sino que son sistemas dinámicos, que poseen propiedades funcionales *emergentes* de la actividad paralela, masiva y distribuida de redes de unidades interconectadas. A estas propiedades emergentes les damos los nombres de «esquemas», «conceptos», «reglas», etc. Las estructuras de conocimiento, en una palabra, son propiedades emergentes de la organización y el funcionamiento masivo y paralelo de esas redes.

(5) Así, los modelos conexionistas se sitúan en un plano más molecular de descripción que los clásicos. Pueden entenderse como «microcognitivos» (Clark, 1989). Desde la perspectiva de tales modelos, las teorías clásicas del paradigma C-R y la psicología del procesamiento de la información, son «metáforas» —útiles por su transparencia semántica— para definir, en términos molares, las propiedades globales que emergen del funcionamiento de las redes. Sin embargo, no serían aceptables los supuestos de que: (a) el sistema cognitivo guía su funcionamiento por reglas sintácticas, (b) que definen y dirigen las transformaciones de expresiones simbólicas, (c) sirviéndose de mecanismos de cómputo serial, más o menos semejantes al imaginado por Turing.

(6) Por otra parte, en los nuevos modelos de la mente no hay nada que se parezca ni al «dispositivo de lectura y control», ni a la cinta con cuadrículas pintadas con símbolos de la máquina de Turing. No hay, en

estos modelos, un «procesador central» con misiones de control. El control es, también, distribuido e implícito, y no localizable y explícito. Se deriva de las propiedades intrínsecas de los sistemas conexionistas —es decir, de las propiedades que se definen en las leyes de propagación de la activación, cambio de los patrones de conectividad o aprendizaje, y activación de las unidades—. Las memorias no se definen por «localizaciones» y sistemas de acceso a ellas, como en los ordenadores de la «familia von Neumann», sino por patrones de conectividad diversos que pueden producirse en un mismo subconjunto de la red.

(7) Por último, y como resulta evidente, los modelos conexionistas no obedecen a la prescripción fundamental de dualismo funcionalista que ha regido la investigación del paradigma C-R y de gran parte de la psicología del procesamiento de la información. La descripción de la mente no puede hacerse con independencia de las propiedades del «hardware» del que es función. Al menos, no puede hacerse sin dejarse en el camino características esenciales de los procesos mentales (flexibilidad, resistencia a la degradación, versatilidad, sensibilidad a la experiencia, etc). Las teorías conexionistas pueden considerarse modelos abstractos del funcionamiento del sistema nervioso; modelos en que se prescinde de aspectos menos relevantes de él (como las transformaciones bioquímicas de las sinapsis, que tanto impresionaban a Searle) para extraer sólo los más pertinentes: aquellos que se refieren a sus propiedades como sistema de cómputo.

Ello no quiere decir que los modelos actuales sean totalmente «realistas» en su representación del sistema nervioso (no lo son, como puede comprobarse al revisar los capítulos 20 a 25 del libro de McClelland, Rumelhart y el grupo de investigación PDP, 1986), ni que puedan establecerse identificaciones simples tales como la de las neuronas con las unidades de los conexionistas (una unidad corresponde, más bien, a conjuntos complejos de neuronas y circuitos, en los modelos hasta ahora desarrollados, que aún son «bocetos» muy generales), pero lo cierto es que las teorías conexionistas están haciendo una contribución decisiva al conocimiento de las propiedades generales de cómputo que, en un nivel abstracto, poseen las redes neurales.

Las observaciones anteriores son sólo algunas de las que pueden utilizarse para demostrar hasta qué punto es exacta la caracterización del enfoque conexionista como aquel en que se acepta el reto de Turing, pero cambiando decisivamente las reglas del juego. De hecho, la Máquina de Turing no es un prototipo tan adecuado de estos modelos como

puedan serlo otras máquinas muy diferentes. Un ejemplo muy revelador, es el de la Máquina de Boltzmann (1986), el autor de la teoría cinética de los gases y creador de la mecánica estadística (vid. Prigogine, 1983). Es una máquina termodinámica, tan abstracta como la de Turing, que posee propiedades intrínsecas de «relajación» de un sistema termodinámico hasta un estado de máxima estabilidad. Estas propiedades se derivan de las funciones estadísticas de las colisiones entre partículas cuyo movimiento y trayectoria no admite ninguna descripción determinista. Del mismo modo, los sistemas conexionistas alcanzan automáticamente estados de «solución óptima» a problemas de reconocimiento de patrones conceptuales o perceptivos a partir de estímulos degradados. Su funcionamiento también puede entenderse como resultante de las propiedades estadísticas de sistemas cooperativos que tienden, de forma intrínseca, a alcanzar estados de «relajación» máxima (en este caso, de acuerdo óptimo con todas las restricciones a las que el sistema se enfrenta). Hinton y Sejnovski (1986) han hecho explícitamente esa comparación entre la máquina de Boltzmann y las redgs de procesamiento distribuido y paralelo que constituyen los nuevos modelos de la mente.

Frente al modelo mecanicista de Newton, basado en supuestos de reversibilidad y determinismo de los procesos físicos, las intuiciones de Boltzmann (1986) se corresponden con la idea de sistemas que son, a un tiempo, estocásticos e irreversibles. Sistemas capaces de *auto organización* de alcanzar niveles de orden cada vez más altos a partir del desorden (Prigogine, 1983). Estos sistemas parecen más coherentes, en muchos aspectos, con intuiciones fundamentales sobre la mente que el autómata simbólico de Turing y, como mínimo, pueden ayudarnos a explicar los mecanismos más minuciosos y moleculares por medio de los cuales ese autómata decerebrado puede transitar de un símbolo a otro.

Así, podemos decir que aceptar el desafío propuesto por Turing, asumir el principio fundamental de toda la ciencia cognitiva de que es posible una descripción computacional y abstracta de lo mental en términos de las *formas* en que lo mental se produce, no significa que tenga que realizarse el juego con las reglas establecidas por el propio Turing en su descripción de la máquina simbólica universal. Hay otras reglas y otras intuiciones que pueden enriquecer enormemente el juego.

Nuestros modelos actuales de la mente son así diversos, y se han beneficiado del rigor conceptual del paradigma C-R, del enorme esfuerzo empírico y de elaboración teórica de la psicología del procesamiento de la información y de las intuiciones y formalizaciones recientes de los

conexionistas. La mente es, a la vez, capaz de conocer (y así puede decirse con propiedad que posee «conceptos», «esquemas», etc) y de autoorganizarse para enfrentarse, de forma cada vez más adaptativa, a un medio cambiante y ruidoso. Actúa de forma «casi-reglada», pero no tan reglada como para ser inflexible. Es función de un organismo, pero no puede entenderse como determinada, en sus patrones, por unarígida adherencia a particularidades minuciosas de él. Posee, en cierto sentido metafórico, una sintaxis, pero no la sigue siempre a ciegas.

Es la mente tan compleja y flexible que sólo podemos entenderla por medio de perspectivas plurales, cuando la miramos desde todas sus esquinas. ¿Seremos así capaces de enfrentarnos a la inquietante posibilidad planteada por Johnson-Laird (1983) de que quizá toda mente sea, a la vez, demasiado simple como para conocerse a sí misma y demasiado compleja como para ser conocida?

Capítulo 5

CIENCIAS COGNITIVAS Y TECNOLOGIAS DEL CONOCIMIENTO

5.1. La(s) ciencia(s) cognitiva(s): ¿Singular o plural?

Desde hace dos décadas, viene formulándose, cada vez con mayor insistencia (Colins, 1977; Gardner, 1985; Mandler, 1981; Norman, 1980; Posner,1989; Pylyshyn, 1984; Rumelhart, 1982;, Varela, 1990) la idea de que existen las premisas necesarias para el desarrollo de una ciencia cognitiva, que incluiría a la Inteligencia Artificial, junto con la Psicología Cognitiva, y partes importantes de la Lingüística y la Lógica. Además, y dependiendo de las especialidades y preferencias teóricas de los que formulan la idea, se incluyen a veces otras ciencias y saberes, tales como sectores de las neurociencias y de la epistemología (Varela, 1990), de la antropología cultural y la filosofía de la mente (Gardner, 1985), etc. Hay un libro reciente, editado por Michael Posner, *Foundations of Cognitive Science* (1989), que constituye un intento, relativamente exhaustivo, de compendiar conceptos, datos y modelos teóricos con los que ya cuenta(n) la(s) nueva(s) ciencia(s).

Como quizá resulten molestos para la lectura los paréntesis y eses que se incluyen al final, cada vez que hablamos de ciencia(s) cognitiva^), es conveniente que aclaremos esa ambigua indecisión entre singular y plural que parecen querer indicar. ¿Debemos hablar de «ciencia» o de «ciencia(s) cognitiva(s)»? El asunto no está claro, y ha dado lugar a un debate no tan intranscendente como parece a primera vista. Explicaré por qué.

La idea de la Ciencia Cognitiva está indisolublemente ligada a la de la formación de una sola categoría (Fm) «mentes/máquinas». Todos los defensores de la tesis de una o varias ciencias cognitivas están explícita-

mente de acuerdo en que la Psicología Cognitiva y la Inteligencia Artificial ocupan puestos centrales en el proyecto. Por ello, la posición de los distintos investigadores, con respecto a la idea de la(s) ciencia(s) cognitiva(s), depende mucho de su posicionamiento con respecto al fundamento, y el posible carácter literal o metafórico, de esa nueva categoría. Es evidente, después de lo que hemos visto en el capítulo anterior, que la propuesta de la ciencia cognitiva no tiene sentido desde la perspectiva de Searle, por ejemplo, y, por el contrario, es muy necesaria para aquellos que defienden una interpretación *literal,* al nivel de abstracción que sea, de la categoría (Fm) que incluye a mentes y máquinas como sistemas de cómputo sobre representaciones simbólicas (i.e. para los defensores de lo que hemos llamado «paradigma C-R»).

Probablemente sea, una vez más, la formulación de Zenon Pylyshyn en *Computación y conocimiento* (un verdadero libro de «ciencia cognitiva») la que expresa de forma más clara este enfoque, que implica la defensa de una ciencia sustantiva, autónoma e integrada: la ciencia cognitiva sería singular; no resultante de una «unión política» de ciencias distintas que dialogan sobre el conocimiento, sino «un auténtico dominio científico, al igual que la química, la biología, la economía o la geología» (ed. esp. 1988, p. 10). Un dominio definido por todos los objetos capaces de conocer, mentes y ordenadores. En esta interpretación, la misión de la ciencia cognitiva sería la de establecer los principios comunes a los seres capaces de conocimiento y computación y, muy en el estilo del paradigma C-R, tal ciencia no sería un mero resultado de la inducción de esos principios a partir de los datos obtenidos por las ciencias particulares. Más bien al contrario, establecería los cimientos de tales ciencias (no sus tejados).

Una posición semejante es la que mantienen Newell y Simón en un artículo importante, que tiene un título muy revelador: «Computer Science as Empirical Enquire: Symbols and Search» (1976; vid. reed. en Boden, 1990). Las ciencias de la computación serían «empíricas» y «cada nueva máquina que se construye es, en sí misma, un experimento» (p. 106). La Ciencia Cognitiva sería, así, la que tendría el objetivo de formular las leyes y principios comunes a todos los sistemas físico-simbólicos, entre los que se incluirían las mentes y los sistemas artificiales de cómputo. Luego las «ciencias empíricas particulares», tales como la psicología y la I.A., establecerían las características específicas de los sistemas naturales y artificiales respectivamente.

Gardner (1985) ha formulado los que podríamos llamar (en la tradición del estudio de los conceptos) «rasgos definitorios», es decir, rasgos que son necesarios para que una pieza de conocimiento se defina como verdadera «ciencia cognitiva» de la ciencia cognitiva, y «rasgos característicos» de ella, que no son imprescindibles pero sí frecuentes. Entre los primeros incluye dos: (1) la ciencia cognitiva postularía un nivel de análisis en que el concepto de «representación» tendría un papel central; un nivel «totalmente separado del nivel neurológico o biológico, por un lado, y del sociológico o cultural, por el otro» (p. 22), y (2) para la comprensión de la mente, el ordenador sería «el modelo más viable» (ibíd.)

Los rasgos característicos (que serían «estrategias», más que principios fundamentales de la ciencia cognitiva) serían estos tres: (1) la tendencia a no tomar en consideración variables emocionales, histórico-culturales, etc, (2) la interdisciplinariedad, y (3) la atención a problemas epistemológicos planteados por la tradición filosófica de Occidente.

Esta caracterización (quizá como cualquier otra que pudiera hacerse) es discutible. Parece claro que la ciencia cognitiva tiene que hablar necesariamente de *representaciones* y *procesos* sobre ellas. No lo es tanto que tenga que situarse en un nivel de discurso totalmente separado del biológico (¿no caben los conexionistas?) y el social (¿tampoco Vygotski, ni la etnociencia, que revisa el propio Gardner?). Después de la revisión breve que hemos hecho de algunas ideas conexionistas, será evidente que tampoco está tan claro que los ordenadores de la «familia von Neumann» sean más viables, como modelos de la mente, que máquinas de tratamiento de la información de estilo Boltzmann, por ejemplo. Por otra parte, los rasgos característicos sí lo son y quizá definan más, actualmente, a la ciencia cognitiva que los considerados definitorios.

Creo que puede hacerse una definición breve de la ciencia cognitiva no tan comprometida con el paradigma C-R como la de Gardner: es la ciencia general de los sistemas capaces de realizar cómputos en función de representaciones. Establece un nivel de discurso que se refiere a las *formas* de éstas y aquéllos, y no a su implementación física. Se sitúa en un plano esencialmente *funcional.* Tiene una clara vocación formalizadora, un interés epistemológico y un carácter fuertemente interdisciplinar.

Para los psicólogos situados en el marco de los modelos de procesamiento de la información, la categoría (Fm) «mentes/máquinas» suele tener (ya lo hemos comentado) una naturaleza metafórica, más que literal. En este enfoque, la ciencia cognitiva toma un carácter más plural: constituye un ámbito de encuentro y diálogo de ciencias particulares que

se intercambian conceptos y datos que pueden tener un valor heurístico. Los psicólogos obtenemos observaciones, y formulamos teorías que son útiles para los investigadores en Inteligencia Artificial, en tanto que éstos traten de desarrollar sistemas con propiedades o ventajas del sistema cognitivo humano. A su vez, los propios psicólogos pueden y deben incorporar modelos de la I.A., y de otras disciplinas como la lingüística teórica y la lógica formal.

Desde esta perspectiva, la ciencia cognitiva tiende a considerarse una posibilidad más lejana, y más diversa, en su configuración, que cuando se concibe desde el supuesto C-R , según el cual la categoría (Fm) «mentes-máquinas» admite una interpretación literal. Con arreglo a la cauta tradición empirista que aún pesa tanto en psicología, muchos teóricos del procesamiento de la información prefieren concebir la ciencia cognitiva como un *resultado,* o una *consecuencia,* más o menos deseable, de la inducción de principios comunes definidos por psicólogos, lógicos, lingüistas e investigadores de la Inteligencia Artificial y las neurociencias. Desde esta perspectiva, contamos, por ahora, con *ciencias cognitivas.* No, quizá, con los elementos suficientes para hablar de una ciencia cognitiva unitaria e integrada.

Los dos axiomas principales de los que parte el proyecto de la Ciencia Cognitiva son, como ya hemos comentado, los siguientes: (1) Las mentes y ciertas máquinas pueden incluirse —literal o metafóricamente— en una sola categoría de seres capaces de conocimiento, y (2) el conocimiento que poseen mentes y máquinas puede explicarse en términos esencialmente *formales* y *funcionales* sin tener que basar la explicación, necesariamente, en la organización material del sustrato biológico, o mecánico, capaz de conocer. Esta segunda suposición otorga al proyecto de una ciencia cognitiva integrada un carácter marcadamente *logicista* (vid. De Vega, 1981). El lema de que «conocer equivale a formalizar», aunque no se formule explícitamente, es asumido por los defensores más entusiastas de la Ciencia Cognitiva en singular. Y ese lema lleva, naturalmente, *a partir de las formalizaciones* (lógicas, lingüísticas y de la teoría de la computación) previamente construidas para comprender el funcionamiento de la mente.

Como señala De Vega (1981), el metapostulado logicista que domina, en una primera fase, la investigación cognitiva en psicología establece, implícitamente, que «las representaciones y/o procesos mentales son isomorfos respecto al sistema de reglas formales lógico o matemático (vg. lógica de proposiciones moderna y lógica de predicados de primer orden)» (p. 347). Este metapostulado ha jugado un papel importante en las primeras formulaciones de modelos cognitivos: en los estudios sobre razonamiento silogístico, las investigaciones clásicas sobre formación de conceptos, las teorías proposicionales de la comprensión y la memoria, las concepciones primeras sobre la influencia de reglas sintácticas en la comprensión, los modelos sobre predicción, etc.

Sin embargo, a medida que los estudios empíricos sobre el funcionamiento mental en tareas de razonamiento, comprensión, predicción, solución de problemas y memoria, han puesto de manifiesto las desviaciones de los procesos y representaciones reales, que emplean los sujetos, con respecto a los modelos normativos de tipo lógico, matemático o lingüístico, se ha producido un proceso de inversión de la dirección de influencia: actualmente se pide a los lógicos, matemáticos y teóricos de la I.A (a los «formalizadores») que desarrollen formalizaciones capaces de simular las conductas reales de los sujetos cuando piensan, comprenden y recuerdan (Riviére, 1987). La formalización de «guiones», lógicas y conjuntos borrosos, estructuras lingüísticas con fuertes implicaciones semánticas o pragmáticas, etc., son buenos ejemplos de ellos.

Marvin Minsky, en *La sociedad de la mente* (1986), un libro lúcido y profundo que introduce, de forma muy clara, a conceptos fundamentales de la ciencia cognitiva, explica que el supuesto primero (mecanicismo) de esta ciencia no tiene por qué llevar al segundo (logicismo):

> Ser capaces de explicar en términos lógicos cómo funcionan las partes de una máquina —dice— no nos permite, automáticamente, explicar sus actividades posteriores en términos simples y lógicos. Edgar Allan Poe sostuvo una vez que cierta «máquina» de jugar al ajedrez debía ser un fraude porque no ganaba siempre. Si verdaderamente fuera una máquina, afirmaba, debería ser perfectamente lógica, ¡y por lo tanto nunca podría cometer errores! ¿Cuál es la falacia aquí? *Simplemente, que nada nos impide emplear un lenguaje lógico para describir un razonamiento ilógico* (p. 192).

La relación entre lógica y psicología, en el ámbito de la ciencia cognitiva, ha empujado a los lógicos a enfrentarse a la tarea que propone Minsky: emplear un lenguaje lógico para describir un razonamiento ilógico.

Así, la psicología empírica del conocimiento ha contribuido a ampliar, de forma decisiva, el marco de definición de las ciencias formales que forman parte de la ciencia cognitiva. Por ejemplo, la lógica se ha entendido tradicionalmente como la ciencia de las inferencias formalmen-

te válidas (Deaño, 1980), pero, en un sentido más general, los lógicos pueden jugar un papel decisivo en la formalización de inferencias que llevan a cabo sistemas naturales o artificiales de procesamiento, con independencia de la validez formal de tales inferencias.

El proyecto de una ciencia cognitiva integrada, o de una «epistemología científica» se ha formulado desde perspectivas relativamente alejadas del paradigma C-R, pero que partían del axioma logicista y de la vocación formalizadora que caracteriza a ese paradigma. Este es el caso de Piaget (1970). En su enfoque, la lógica establecería una axiomática de la razón, cuya ciencia experimental sería la psicología de la inteligencia. La respuesta a cuestiones epistemológicas tradicionales adquiriría un carácter científico al situarse entre esas dos fronteras, la de la lógica y la de la psicología de la inteligencia, definiendo, por procedimientos empíricos y formalización estructural, la psicogénesis de las estructuras de conocimiento. Además, otros saberes rigurosos —la historia de la ciencia, la sociología del conocimiento, etc— se integrarían en ese proyecto común de una ciencia integrada del conocimiento que daría respuestas a viejas preocupaciones epistemológicas. Como es sabido, Piaget no se limitó a formular teóricamente la posibilidad de esa epistemología científica, sino que, durante muchos años, se preocupó de darle una cobertura organizativa a través de las reuniones interdisciplinares en el Centro de Epistemología Genética de Ginebra.

Tanto en la perspectiva piagetiana, como en la de los teóricos del paradigma C-R, la psicología del conocimiento cumple una función central e insustituible en la ciencia del conocimiento o la epistemología científica. Ese lugar le corresponde en tanto que es la ciencia empírica de los sistemas que son, por antonomasia, capaces de conocer: las mentes. Trabaja sobre los únicos objetos de cuya capacidad de conocimiento no existen dudas: los organismos.

Siguiendo la distinción, trazada en otro momento, entre el componente sintáctico y el semántico de las teorías científicas, podemos decir que la psicología cognitiva es esencial para la definición de la base semántica de la ciencia cognitiva globalmente considerada. Su función es decisiva para establecer la base observacional de cualquier intento de conocer ese objeto tan complejo y elusivo que es el propio conocimiento.

5.2. Un enfoque nuevo de viejos problemas epistemológicos. Psicología cognitiva y tecnologías del conocimiento

El intento de explicar el conocimiento no es nuevo. Ha sido una obsesión tenaz en los últimos dos mil quinientos años, y seguramente se remonta a los orígenes del hombre como especie. Sin embargo, la psicología cognitiva en particular, y la ciencia cognitiva más en general no son simples prolongaciones de las viejas preocupaciones epistemológicas que ya manifestaron los griegos de los siglos VI y V a C. Implican un enfoque fundamentalmente nuevo, que sólo ha sido posible a partir de la ruptura con la distinción entre mentes y máquinas en que se basó el nacimiento de la ciencia desde Kepler, Copérnico, Galileo y Newton.

En un librito reciente sobre las tendencias y perspectivas de las ciencias cognitivas, titulado *Conocer,* Francisco Varela (1990) da una explicación muy sintética y elegante del que es, sin duda, el núcleo esencial de novedad de estas ciencias. Las ideas científicas acerca de la naturaleza, dice Varela, forman parte de una cierta constelación cognitiva o «estructura imaginaria» que produce cada época a partir de sus prácticas sociales y culturales y de su lenguaje: «los historiadores y filósofos modernos, desde Alexander Koiré, han demostrado que la imaginación científica sufre mutaciones radicales de una época a otra, y que la ciencia se parece más a una epopeya novelística que a un progreso lineal. La historia humana de la naturaleza es una narración que merece ser contada de más de un modo» (ibíd., p. 12).

Esas concepciones de la naturaleza —así como las propias prácticas sociales y desarrollos técnicos, deberíamos añadir— se corresponden con las ideas sobre el sí mismo, sobre la mente, que las distintas épocas producen, de tal manera que «reflejándose mutuamente, el sí mismo y la naturaleza se desplazan en el tiempo como una pareja de bailarines» (ibídem). Pues bien: las ciencias cognitivas representan un cambio importante en esa historia paralela de mente y naturaleza, una alteración esencial en ese baile. Lo que ha sucedido es que, por primera vez en la historia, el enfrentamiento al problema del conocimiento ha podido ligarse al desarrollo de una tecnología que está transformando, a la vez, las prácticas sociales y los conceptos de la mente: la tecnología del conocimiento cuyo aspecto más visible —y más influyente en psicología— es la Inteligencia Artificial.

Varela advierte, con toda razón, que «no podemos separar las ciencias cognitivas de la tecnología cognitiva sin despojar a una u otra de un

vital elemento complementario». La tecnología cognitiva propone a la mente una nueva imagen de sí misma. Al mismo tiempo, espolea su pretensión de conocerse y define algunos de los parámetros a que debería atenerse ese conocimiento. La imagen es la que corresponde a un mecanismo de cómputo que transforma representaciones, y que discurre por debajo de esas otras imágenes aparentes que se ofrecen a la visión de la conciencia. Esa imagen está indisolublemente ligada a la impresión, realista o ilusoria, que el hombre tiene, en la segunda mitad de nuestro siglo, de que está produciendo con sus técnicas rasgos y atributos que se han predicado siempre, de forma intuitiva, de lo que nosotros hemos llamado «categoría natural Fm».

Como veremos en el capítulo próximo, un rasgo esencial de los objetos a los que se atribuyen funciones Fm es el hecho de que, como decía Descartes en un párrafo citado anteriormente, *tales objetos obran por conocimiento* y no, de manera refleja, por la «simple disposición de sus órganos». Su forma de comportarse depende, en otras palabras, de su capacidad de «contener» el medio en forma de representaciones internas, y de transformar tales representaciones haciéndose capaces de anticiparse a modificaciones previsibles en el medio. En el hombre, estas capacidades llegan a su máxima expresión cuando se hace capaz de «comprender lo real como un subconjunto de lo posible», como señala Piaget. Hasta la segunda mitad de nuestro siglo parecía que sólo los organismos poseían esa capacidad de guiar su conducta por representaciones internas —esa competencia, dicho de otro modo, de tener una mente intencional—. El desafío de Turing no es más que un indicio de que esa situación ha cambiado fundamentalmente en los últimos cincuenta años.

De este modo, la revolución cognitiva (que no sólo afecta a la psicología, sino a todas las ciencias que forman parte del círculo de ciencias cognitivas al que nos referíamos en el apartado anterior) puede considerarse, desde una cierta perspectiva, como la segunda gran revolución mecanicista de la historia de la ciencia. La primera fue la que se produjo en los siglos XVI y XVII e implicó, gracias a la obra de los físicos del Renacimiento y el Barroco, la reducción mecanicista del mundo lejano e inaccesible del macrocosmos. La segunda, en los últimos años, ha supuesto la reducción mecanicista de un mundo más cercano e íntimo, pero no más accesible: el mundo de lo mental.

Una condición importante para que se produjera esta segunda revolución mecanicista fue la transformación y ampliación del concepto mismo de «máquina». La concepción determinista, causalista, y basada en

el supuesto de reversibilidad de la mecánica clásica, no es válida para explicar los fenómenos termodinámicos ni el funcionamiento de las máquinas térmicas. Estos fenómenos y máquinas, regidos por un principio de crecimiento de entropía que establece una «flecha» irreversible en el tiempo, se definen por una descripción que ya no consiste —como en la cinemática clásica— en una definición de los movimientos y estados de equilibrio de masas rígidas, sino en la definición de magnitudes macroscópicas que dependen del sistema en su totalidad (Aracil, 1986). El paso de las «máquinas mecánicas» a las «máquinas termodinámicas» fue el primero que permitió una ampliación decisiva de la mecánica, sobre todo gracias a la obra de Boltzmann (1986).

El segundo momento importante de esta transformación es el definido por la creación de máquinas que exigen un componente de *información* para funcionar. Las primeras fueron, precisamente, máquinas termodinámicas. Uno de los ejemplos más característicos es el del regulador centrífugo de Watt, que permite ajustar la velocidad desarrollada por una máquina de vapor a su carga, mediante un sistema de realimentación negativa (al aumentar la velocidad, unas pesas se separan de su eje y este movimiento cierra un anillo en el sistema de inyección de vapor a una caldera). Estas máquinas, capaces de autorregulación en función de una realimentación negativa, se hacen cada vez más complejas en la medida en que —como en el caso de los cañones antiaéreos automatizados— se construyen de manera que puedan anticipar, mediante cálculos de carácter estadístico, valores predichos de la variable a que tienen que «adaptarse», y de corregir de forma gradual desviaciones cada vez menores con relación a esa variable.

La posibilidad de incorporar mecanismos de procesamiento de la información en las máquinas —dice Aracil— permite obtener de ellas formas de comportamiento que ponen de manifiesto enormes dosis de habilidad e ingeniosidad, y que están claramente separadas de los comportamientos de tipo mecanicista, de carácter esencialmente repetitivo, de los juguetes mecánicos y de los relojes. La introducción de la información en la concepción de las máquinas conlleva la incorporación en su estructura de mecanismos de realimentación, que están en los orígenes de la cibernética. Gracias a estos mecanismos las máquinas muestran formas de comportamiento en las que ponen de manifiesto aspectos adaptativos y aparentemente teleológicos, sensiblemente diferentes a los mecanicistas (1986, p. 67).

Así, del mismo modo que un ordenador de tipo von Neumann, que

incorpora una especificación de la máquina de Turing, pasa con éxito el «test objetivo» de pensamiento del propio Turing, el cañón antiaéreo que se anticipa delicadamente a la trayectoria predicha de su blanco simula la fina regulación y el carácter anticipatorio de las conductas propositivas a las que solemos atribuir un origen mental. El primero que vio las posibilidades de una síntesis entre mecanismos orgánicos —desde la homeostasis del azúcar a los procesos mentales— y sistemas artificiales autorregulados por realimentación negativa fue el creador de la cibernética, Norbert Wiener, que intentó dar al concepto de propósito un carácter científico, despojándolo de adherencias teleológicas superficiales. Wiener también veía una similitud funcional —definida en la estructura de los bucles de realimentación— entre las conductas propositivas complejas, que queremos estudiar los psicólogos cognitivos, y las que muestran los sistemas mecánicos que corrigen su comportamiento en función de una información o, como diría Descartes, «obrando por conocimiento».

Los primeros psicólogos cognitivos adoptaron las intuiciones cibernéticas para formular conceptos con los-que dar cuenta de las funciones mentales más complejas, e incluso de las conductas más simples. Así, Miller, Galanter y Pribram (1960) proponen una nueva unidad de análisis de la conducta, el TOTE (Test-Operator-Test-Exit), que no es más que un bucle de realimentación negativa. Estas unidades TOTE se encajarían unas en otras, y configurarían las estructuras subyacentes a una amplísima gama de conductas que irían desde reflejos simples a las actividades planificadas y propositivas más elaboradas.

El funcionamiento de sistemas muy complejos, como los ordenadores de tipo von Neumann, en tareas de solución de problemas, puede entenderse también como guiado frecuentemente por estrategias de realimentación, que son unidades TOTEs con un alto grado de elaboración, encaminadas a reducir progresivamente la discrepancia entre un estado inicial (el marco definido simbólicamente por el «espacio de problema» primero) y uno final (de solución). Han sido Newell y Simón (1972, 1976) los investigadores que han definido, de forma más precisa, ese principio estratégico fundamental, que sería común a mentes y máquinas cibernéticas guiadas por representaciones simbólicas, tales como algunos programas de ordenador. Para Newell y Simón (1976), la «Hipótesis de búsqueda heurística», que define de manera global la estrategia autorreguladora de los mecanismos de solución de problemas, es la segunda ley de estructura cualitativa de la mente (recordemos que la primera era la

que establecía que los sistemas físico-simbólicos son la condición necesaria y suficiente de la inteligencia).

Esa ley de estructura cualitativa (que, repitámoslo, no es más que una elaboración sofisticada del mecanismo cibernético de autocorrección por realimentación negativa) establece lo siguiente:

Hipótesis de la búsqueda heurística . Las soluciones a los problemas se representan como estructuras simbólicas. Un sistema físico-simbólico ejerce su inteligencia en la solución de problemas por búsqueda— es decir, generando y modificando progresivamente estructuras de símbolos hasta producir la estructura de solución (1990, p. 119).

A diferencia de los mecanismos cibernéticos simples, como los de los termostatos o los de la homeostasis de la glucosa, que no son universales y tienen una función limitada de adaptación a un solo parámetro o un conjunto muy limitado de ellos, los sistemas mentales o artificiales de solución de problemas se enfrentan, con frecuencia, a medios en que la información potencialmente relevante proviene de fuentes muy diversas y crece de forma exponencial, y deben enfrentarse a esos medios con *recursos limitados.* En esas situaciones, el cumplimiento de sus propósitos, de anulación de la discrepancia entre condiciones iniciales y finales, exige una refinada elaboración de mecanismos de reducción de los espacios y los algoritmos de búsqueda.

Un buen ejemplo es el de los programas de ajedrez. Su estado inicial está claramente definido por la igualdad de fuerzas en el tablero entre el programa y su oponente humano. Todo el programa puede concebirse como un enorme y elaboradísimo sistema TOTE, que trata de pasar de ese estado inicial de igualdad de fuerzas a otro final, definido por un jaque mate al rey contrario. En cada movimiento, el programa se servirá de subsistemas TOTE para alcanzar propósitos parciales de acrecentar ventajas posicionales propias, disminuir recursos ajenos, etc. Pero tales subsistemas no se enfrentan ahora a una tarea tan simple y definida por un conjunto muy limitado de valores de dirección y distancia como en el ejemplo de Miller, Galanter y Pribram del sistema TOTE para clavar un clavo en la pared. Por el contrario, se enfrentan a árboles de búsqueda que pueden tener más de un millón de ramas. En estas ramas —o en algunas de ellas— el propósito de la búsqueda no es generar soluciones sino *evaluarlas,* como en los mecanismos TOTE.

La mente que se enfrenta al programa se encuentra ante la misma situación. También puede entenderse su funcionamiento, a grandes ras-

gos, por el juego de estructuras TOTE, que antes de mover evalúan consecuencias de jugadas posibles. Sabemos, por investigaciones empíricas, que la estrategia que emplean los buenos jugadores consiste en buscar, de forma muy selectiva y relativamente profunda en unos pocos nodos de los posibles. Los grandes maestros no suelen explorar árboles de más de un centenar de ramas. Se concentran en un subconjunto muy limitado del formado por los miles de ramas posibles, y analizan con cierta profundidad ese subconjunto. Como señalan Newell y Simón (en el artículo de 1976, recogido en Boden, 1990), «esto sólo es posible, sin deteriorar las evaluaciones, introduciendo mecanismos de selección en el propio generador, de forma que sea capaz de generar sólo aquellas ramas que muy probablemente brindarán información relevante acerca de la posición» (p. 124). Esos mecanismos pueden implicar el empleo no-local de la información (como en el programa HEARSAY de reconocimiento del habla), el uso de un conocimiento semántico rico sobre el campo de interés (por ejemplo, los grandes maestros de ajedrez almacenan una amplia gama de posiciones y «patrones» generales que admiten soluciones semejantes) o la selección de representaciones apropiadas (Newell y Simón, 1976; Riviére, 1986), etc.

Las máquinas que conjugan una poderosa capacidad de tratar conocimiento, simbólicamente representado, con la competencia de autorregulación propia de los sistemas cibernéticos, producen una clara y definida impresión intuitiva de poseer atributos mentales y desdibujan la distinción mentes (Fm)/máquinas(-Fm).

Creo que un ejemplo que registra una experiencia muy común, puede servir para ilustrar la fuerza y el sentido de la atribución de rasgos Fm a ese tipo de máquinas. El ejemplo se refiere al mismo campo que nos ha servido para mostrar la importancia de las estrategias de búsqueda en los sistemas que se autorregulan por estructuras complejas de conocimiento: el ajedrez.

Basta con cambiar el programa de la máquina con la que se escriben estas páginas, para poder librar con ella encarnizadas batallas de ajedrez. En principio, podríamos pensar en tres estrategias diferentes para jugar: basarse en la constitución física de la máquina (en el *hardware),* basarse en su *software* (en un conocimiento de su programa de ajedrez tal que nos permitiera saber, por ejemplo, que almacena un cierto conocimiento «semántico» de aperturas, o que selecciona ramas de análisis en función de la respuesta a preguntas definidas como: «¿Está en riesgo alguna pieza importante?», «¿Puede ser atacado el rey en la próxima?»), o finalmente basarse en *estrategias, propósitos, conocimientos* y una cierta *racionalidad* a la máquina. En pocas palabras: las tres estrategias posibles serían: (a) la del hardware, (2) la del software y (3) la de lo que los filósofos han llamado *actitudes proposicionales* (tales como los deseos, creencias y propósitos). Nuestro ejemplo puede servir para poner de manifiesto tres consideraciones importantes:

(1) La diferencia entre el lenguaje máquina (más directamente relacionado con el hardware), el lenguaje de programación y el lenguaje de las actitudes proposicionales se produce, en primer término, en una dimensión de molecularismo-molarismo. Cada expresión del lenguaje de programación es interpretable en términos de numerosas expresiones del lenguaje máquina. A su vez, cada expresión del lenguaje de las actitudes proposicionales puede traducirse a un conjunto amplio de expresiones del lenguaje de programación. Así, hay un orden «L.M-L.P-A.P» que va de lo más molecular a lo más molar.

(2) Sólo en los niveles de lenguaje de programación y actitudes proposicionales aparecen expresiones simbólicas, en el sentido de «semánticamente transparentes». Cuando se realiza el programa de ajedrez, el lenguaje máquina *representa,* en cierto modo, jugadas, posiciones, etc, pero lo hace de forma *semánticamente opaca.*

(3) El lenguaje de las actitudes proposicionales es el único que resulta realmente útil para vencer en la partida de ajedrez a la máquina y «comprender», en un tiempo aceptable y sin excesiva carga de procesamiento, lo que ésta hace. En suma, el lenguaje intencionalista, compuesto de verbos de creencia y deseo, de la psicología cotidiana (Dennett, 1978; López Cerezo, 1989), es el lenguaje con el que nos enfrentamos a la máquina.

Este último lenguaje es el que efectivamente usa el jugador de ajedrez cuando juega contra el programa de ajedrez. Sin darse mucha cuenta de ello, la atribuye conocimientos, propósitos y metas, intenciones y deseos. Se ve a sí mismo diciéndose cosas como estas: «¡Ajá!...lo que *quiere* ahora es encerrarme la reina con el caballo y el alfil», o «voy a hacer que crea que ataco el peón de rey, para abrir camino al ataque al enroque», etc. Hay personas que se *indignan* con los programas y les atribuyen trampas y malas intenciones. Es verdad que estas atribuciones no son tan consistentes como las que se hacen con las personas y, sobre todo cuando se pierde, suelen deshacerse rápidamente («¡Bah!...al fin y al cabo es una máquina, ¿a mí que más me da?»). Pero, en los momentos más calientes de la partida, el jugador se sorprende a sí mismo tratando

a la máquina como a un sistema intencional, y poniendo toda su «honrilla» en el empeño intencional de vencer a ese otro sistema intencional, cuya engañosa «voluntad de victoria» se opone inteligentemente a la suya.

Este ejemplo es útil porque permite entender cómo las *actitudes proposicionales,* que definen el vocabulario mentalista de la psicología cotidiana, pueden entenderse como: (a) artefactos conceptuales prácticos, que sirven para comprender y predecir la conducta de sistemas autorregulados complejos «que obran por conocimiento» y (b) que hacen referencia a propiedades molares y emergentes del funcionamiento de tales sistemas (propiedades globales a las que llamamos propósitos, creencias, pensamientos, etc). El lenguaje de las actitudes preposicionales, por otra parte, no sólo resulta útil para la comprensión y predicción de la conducta de ciertos organismos complejos, sino también —como pone de manifiesto el ejemplo anterior— de ciertos sistemas artificiales de procesamiento de la información.

En comparación con la mente, estos sistemas tienen además una gran ventaja para el psicólogo: las máquinas son sistemas cuya estructura y funcionamiento se conoce de forma completa, desde el momento en que son productos del ingenio humano. Sus pasos en el tratamiento del conocimiento, sus estrategias de búsqueda y solución de problemas, sus algoritmos en una palabra, están perfectamente definidos. Si podemos conocer tan bien la organización interna, las estructuras simbólicas, y los procesos de transformación de conocimiento que se producen en esos sistemas mecánicos que aparentan intencionalidad e inteligencia, ¿no podremos comprender igual de bien un sistema, como la mente, que es «realmente» intencional e inteligente?

Esta cuestión renovó, al comenzar la segunda mitad de nuestro siglo, el interés por conocer la mente, y, como comentábamos antes, estableció criterios nuevos de precisión algorítmica y rigor conceptual para ese conocimiento.

La irrupción de la tecnología del conocimiento, con sus nuevas metáforas o sus nuevos conceptos literales —en la interpretación C-R— sobre lo mental, tuvo lugar en un momento en que el interés de la mente por alcanzar un conocimiento científico de sí misma había decaído visiblemente en psicología. El predominio conductista fue una manifestación de ese decaimiento. Los conceptos tradicionales sobre la mente no habían sido fáciles de encajar en esquemas explicativos científicos que fueran más allá de las asociaciones simples del empirismo clásico (criticadas ya por James, 1890, y por el propio Wundt) o de las categorías puramente descriptivas de los psicólogos de la forma (Kohler, ed. 1972). La metodología clásica de conocimiento de la mente, la introspección, estaba agotada después de las polémicas sobre el pensamiento sin imágenes entre Wundt y los psicólogos de Wurzburgo (Humphrey, 1973). En estas condiciones, la renuncia conductista al estudio de lo mental era bastante semejante al gesto dolido y despreciativo con el que el niño dice que «ya no quiere» el objeto que se le niega.

Esta fue la situación histórica en que aparecieron sistemas autorregulados y guiados por representaciones simbólicas que, al tiempo que presentaban inquietantes semejanzas con los sistemas mentales, admitían una descripción completa y un conocimiento exhaustivo por ser productos tecnológicos de la propia mente humana. Esos sistemas se acompañaron de una densa red de reflexiones sobre lenguajes formales y teoría de la computación, y de conceptos tales como sistema, realimentación, máquina de Turing, sistemas de producciones, procedimientos efectivos y cálculo preposicional, que ofrecían nuevas perspectivas para un conocimiento riguroso de la mente, Así, en la segunda mitad de nuestro siglo, los psicólogos encontraron nuevas herramientas conceptuales y nuevas razones para enfrentarse a su vieja pretensión de conocer la mente.

Es en este sentido en el que decía, en otra ocasión, que «la Psicología Cognitiva es, en realidad, una de las manifestaciones más claras y genuinas del *Zeitgeist* científico, la organización tecnológica y ciertos intereses productivos dominantes en las sociedades tecnológicamente más avanzadas de la segunda mitad de nuestro siglo. Es expresión /.../ de una compulsión hacia la información, la computación y la representación que tiene un significado mucho más profundo e influyente que el de un cambio de paradigmas en una ciencia particular» (Rivière, 1987, p. 16). Varela expresa esta misma idea de otra forma: «por primera vez —dice- la sociedad occidental en su conjunto comienza a enfrentar en sus prácticas preguntas como: "¿Es la mente una manipulación de símbolos? ¿Puede una máquina comprender el lenguaje?" Se trata de preocupaciones que afectan a la vida de la gente y no se limitan a ser teóricas...mientras durante milenios las gentes tuvieron una comprensión espontánea de sí mismas, según la cultura de su época, por primera vez esta visión popular de la mente entra en contacto con la ciencia y es transformada por ella» (1990, p. 14).

La segunda revolución mecanicista, que ha implicado la disolución parcial de la cuidadosa reducción de las mentes a los organismos lograda en una dura oposición a las tendencias animistas del pensamiento, ha

significado una irrupción poderosa de lo que Imre Lakatos (1974) denomina «historia externa» en el transcurrir algo marginal de una ciencia joven como la psicología. La historia externa no admite sólo una reconstrucción racional en términos de la lógica «endógena» de desarrollo de la propia psicología. Hace intervenir en ese desarrollo, de forma decisiva, valores y necesidades sociales, influencias culturales, fuerzas económicas. También factores escasamente racionales, como los relacionados con la imagen y consideración social que se tiene de una ciencia, la organización académica y de investigación que la sustenta, los mecanismos que determinan la selección de las personas que se dedican a ella y la distribución económica de los recursos de investigación.

Cualquier intento de explicar el origen y desarrollo de la Psicología Cognitiva en términos exclusivamente internos, y dependientes de la racionalidad intrínseca de la evolución de los conceptos sobre la mente, resulta insuficiente. La otra cara de la historia, a la que nos hemos referido en las páginas anteriores, también tiene que ser reconstruida. El significado de la revolución cognitiva en psicología no se limita al de una revuelta interna y estrictamente racional en el pequeño ámbito de las especulaciones académicas. La Psicología constituye una pieza esencial para el desarrollo de la ciencia cognitiva, y ésta a su vez proporciona un fundamento necesario para la tecnología del conocimiento. Se ha dicho que las ciencias del conocimiento forman parte medular de la tercera gran revolución en la comprensión conceptual del mundo: la que añade un componente de información —o incluso de conocimiento— a una realidad que antes sólo era entendida bien en términos exclusivamente mecánicos o por medio de categorías relacionadas con el concepto de energía. Esto ya lo intuía con claridad Norbert Wiener (1960) cuando relacionaba el concepto de información con el de entropía, una relación que sigue siendo fecunda para el desarrollo de la reflexión sobre el mundo físico (Prigogine, 1986). Por su penetración e impacto en el tejido social, no es exagerada la afirmación de Varela (1990) de que el conjunto compuesto por las ciencias y las tecnologías del conocimiento ha dado lugar a la revolución conceptual y tecnológica más importante de nuestro siglo después de la física atómica.

Estos hechos históricos admiten, como hemos visto, valoraciones muy diversas. Para algunos (Dreyfus, 1979; Searle, 1985, Flores y Winograd, 1989), la creciente presión de la historia externa y la implicación de la psicología en el desafío formulado por Turing, pueden ser negativos a largo plazo: dar lugar a un proceso de pérdida de dirección y sentido.

Lo mental se escaparía por las rendijas explicativas de la metáfora de las máquinas. Para otros —los defensores del paradigma C-R— el concepto de cómputo es el único camino para desarrollar una ciencia de la mente objetiva y respetuosa con la complejidad de lo mental. La psicología, en esta perspectiva, debe incorporarse sin indecisiones, al proyecto integrado de una sola ciencia cognitiva.

Hay un terrible poema de Alain Bosquet, titulado «Todo objeto se vuelve maestro de aquel que lo inventa», que refleja, de forma muy expresiva, la tensión entre esas dos concepciones de la relación entre mentes y máquinas:

> Todo objeto se vuelve maestro
> de aquel que lo inventa,
> y el robot construido por el hombre
> en el universo de lo improbable
> precede al hombre /.../
> El robot piensa
> lo que el hombre no es capaz de pensar:
> es duro, es justo, perfeccionará
> a su inventor, dándole aletas,
> alas, algunos vientres de recambio,
> un doble corazón, un cerebro menos nervioso.
> El robot destruye al hombre:
> ya no tiene edad
> de arreglar sus viejos juguetes.

Las tecnologías del conocimiento no sólo suscitan, de hecho, imágenes sociales positivas sino también valoraciones negativas bañadas de miedo a lo desconocido y a un poder que puede irse de las manos. En la comunidad de los psicólogos también existe una gran diversidad de actitudes confesadas o inconfesas hacia el mecanicismo abstracto que ha servido de base a la construcción de la psicología cognitiva. Pero, con independencia de las actitudes y valoraciones, el hecho histórico verificable es que, sin la influencia de la tecnología del conocimiento, no puede comprenderse el desarrolló de la psicología cognitiva. En un estudio sobre las metáforas de lo mental en los últimos noventa años de historia de la psicología, Gentner y Grudin (1985) (ver también Leary, 1990), a partir de un análisis de las metáforas aparecidas en artículos de *Psychological Review,* han demostrado que las metáforas empleadas por los psicólogos han cambiado radicalmente en un siglo de desarrollo de

la psicología y que, desde los años sesenta, el empleo de metáforas relacionadas con sistemas de cómputo es absolutamente predominante.

Aunque, como dicen Gentner y Grudin (1985), el empleo de estas metáforas no garantice el rigor conceptual de los análisis ni el interés genuino de los modelos teóricos, y sirva algunas veces para revestir la vaguedad con el lenguaje de la precisión, lo cierto es que expresan el hecho histórico profundo señalado por Bosquet: «Todo objeto se vuelve maestro de aquel que lo inventa». A partir de la desafiante inclusión en una misma categoría de «informávoros» de todas las mentes y algunas máquinas, el ordenador ha sido realmente maestro de nuestras concepciones de la mente.

Sección tercera:
LAS FUNCIONES MENTALES

Capítulo 6

PSICOLOGIA NATURAL Y ENFOQUES COGNITIVOS DE LAS FUNCIONES MENTALES

6.1. Conocimiento y comprensión natural de la mente

El análisis de los objetos a los que se atribuyen funciones mentales nos ha permitido desvelar aspectos importantes de los fundamentos de la psicología cognitiva, a partir de la provocativa propuesta de Turing (1950) de considerar las consecuencias conceptuales de incluir autómatas con capacidades universales de cómputo en la categoría (Fm). Al examinar el significado de esa propuesta, y las diferentes respuestas que se han dado a ella, se ponían de manifiesto las importantes influencias de la historia externa en la configuración de la psicología cognitiva en particular, y de la ciencia cognitiva como un todo.

Sin embargo, en ese análisis de los fundamentos de nuestra ciencia, la reflexión sobre la definición extensiva de la categoría (Fm) —es decir, sobre los objetos a los que se atribuyen funciones mentales— no resulta suficiente: debe acompañarse de un examen del aspecto intensivo de la categoría, de la naturaleza misma de las funciones que se incluyen en ella. A ese análisis, que nos permite penetrar de forma más directa en la historia *interna* de la psicología como ciencia de la mente, le dedicaremos esta sección. Seguiremos en ella una línea de exposición paralela a la de la sección anterior: veremos primero las características y propiedades que se asignan a las funciones mentales en los actos cotidianos de pensamiento y lenguaje; y revisaremos algunas propiedades esenciales de lo mental, establecidas por psicólogos de clara propensión filosófica (Brentano, 1874) o bien por filósofos de la mente (Searle, 1983). En el capítulo séptimo, reflexionaremos sobre la significación del hecho mismo de que los seres humanos seamos capaces de atribuirnos funciones men-

tales unos a otros. Finalmente, dedicaremos los tres últimos capítulos de la sección al análisis de los problemas, que se han dado históricamente, para un tratamiento científico del lenguaje mental.

Empecemos por lo más obvio, por algo que hacemos probablemente centenares de veces todos los días: lo que hacemos es sencillamente hablar con nosotros mismos y con los demás y, al hacerlo, nos atribuimos y les atribuimos mentes que tienen las capacidades de conocer, creer, desear, recordar, etc. La mente, aunque no fuera otra cosa, es algo que se *piensa* y *se predica, mediante el lenguaje,* acerca de uno mismo y de los demás. Independientemente de que predicados tales como «pensar que (x)»,«recordar (y)», «creer (z)», «desear (v)», puedan o no reducirse a otros de naturaleza muy diferente tales como «estar en el estado (x) del sistema nervioso», «estar en la relación funcional (y) con el medio», o «presentar un conjunto de activaciones (z) y un patrón de conectividad (p) en una red de procesamiento distribuido y paralelo», lo cierto es que los primeros enunciados («pensar que (x)», «recordar (y)», etc) son los que empleamos en el pensamiento y el lenguaje cotidianos para manejarnos en nuestras interacciones con nosotros mismos y con los demás. Tales enunciados contienen un vocabulario que se compone de lo que los filósofos denominan *actitudes proposicionales,* que incluyen dos elementos: (1) una cierta relación, o acción *interna,* o actitud —en sentido amplio—, que un *sujeto* mantiene con respecto a (2) un cierto *contenido.* Así, en «pensar que (x)», hay que distinguir el modo de relación, «pensar que» del contenido al que apunta, (x); en «recordar (y)», se predica un modo de relación (o una acción interna) diferente de otro contenido, (y), y así sucesivamente.

De tal manera, nuestro vocabulario acerca de nosotros mismos y acerca de los otros se compone muy esencialmente de ciertas estructuras conceptuales y lingüísticas —las actitudes proposicionales— que, en un análisis primero, presentan algunas características peculiares que merecerían un análisis más detallado que el que podemos hacer aquí y que los que se han hecho anteriormente. Debemos referirnos muy brevemente a algunas de esas características:

(1) En el lenguaje cotidiano —y dejando de lado las formas metafóricas— las oraciones que contienen el vocabulario de las actitudes proposicionales alcanzan un grado máximo de adecuación semántica cuando contienen un sujeto con los rasgos de ser animado y humano. Como ya comentábamos en el capítulo segundo, de los sujetos animados/humanos pueden predicarse *todos* los elementos que forman el conjunto de las

actitudes proposicionales. En la medida en que los sujetos se alejan de los rasgos anteriores, la predicación de actitudes proposicionales comienza a ser más selectiva y a resultar semánticamente menos natural, como en «mi perro supone que (x)» o, más aún, «la rana cree que (y)». Cuando los sujetos no poseen el rasgo de «animados», la atribución de actitudes proposicionales resulta semánticamente muy anómala o tiene un contenido claramente metafórico, como cuando Aleixandre dice: «La soledad del mineral es sólo un pensamiento. Pero sin el hombre no vive». La excepción, que hemos comentado extensamente en el capítulo anterior, es «El ordenador piensa que (x)», que admite una interpretación científica literal o un carácter de metáfora científica, aunque siga resultando semánticamente extraña en el lenguaje cotidiano.

(2) Existen variaciones muy sutiles en relación con el *nivel de demanda semántica* de los diferentes verbos mentales o de las distintas actitudes proposicionales. En el nivel más alto de demanda semántica se sitúan las relacionadas con el pensamiento y el lenguaje, que exigen los tres rasgos «animado/animal/humano» en el sujeto. Como veremos en otro momento, estas formas de relación de un sujeto con ciertos contenidos proposicionales poseen la propiedad lógica de ser *intensionales,* y de serlo de forma muy clara. También la propiedad psicológica de ser actividades esencialmente simbólicas.

En términos generales, y de forma rápida, puede establecerse una distinción entre los verbos mentales relacionados con «pensar-decir-comprender» que exigen los rasgos «animado/animal/humano» en el sujeto, y los relacionados con «percibir» y «desear», que generalmente sólo presuponen, en sus sujetos, la presencia de los rasgos «animado/animal», en los usos cotidianos del lenguaje. Aunque el tema ha dado lugar a algunos debates filosóficos (vid., por ejemplo, Searle, 1983), podemos señalar que, en rigor sólo de los primeros puede decirse que definen *actitudes proposicionales,* en el sentido de que «piden» proposiciones como «objetos directos». Que las estructuras resultantes de las actividades de la percepción admitan una definición proposicional en sentido estricto es más bien dudoso. Parece intuitivamente sensata la hipótesis de que los mecanismos perceptivos poseen la función básica de configurar un mundo de *objetos,* y no propiamente de proposiciones o creencias, con independencia de lo «penetrables» que sean los sistemas perceptivos a esas otras entidades más complejas, de las que puede predicarse la verdad o falsedad.

Por su parte, los verbos de deseo y recuerdo parecen ocupar una

posición intermedia en que *es opcional* la elección entre una proposición o un objeto como objetos directos: es posible «desear que (x)» o simplemente «desear algo», «recordar que (y)» o «recordar (algo)». Lo que no parece posible, en rigor, es «pensar algo» sin que sea algo lo que se piensa acerca de ello. De modo que puede decirse que el pensar y el decir, cuando éste no tiene simplemente una función expresiva o imperativa (i.e. cuando es expresión de un sistema de *fijación de creencias,* como es el pensamiento), son actividades que contienen, en sí mismas, una implicación proposicional. Tales actividades, simbólicas y con un compromiso proposicional, son las que se traducen en verbos más exigentes con respecto a los atributos semánticos de los sujetos de los que se predican.

(3) Los elementos conceptuales básicos de los enunciados que predican actitudes proposicionales son los verbos mentales. Si bien no se ha realizado aún un análisis semántico completo de tales verbos, Searle (1983) ha especulado con la hipótesis de que podrían descomponerse, en último término, en elementos semánticos de creencia/deseo (así «temer que p» sería equivalente a «creer que es posible que p» y «desear que no-p»: (Temer (p) = Creer (p) & Des (-p)), y «sentirse defraudado de p» equilvadría a (S. defr (p) = pres Creer (p) & pas Creer (fut -p) & Des (-p))), en un lenguaje lógico modal y con cláusulas temporales.

Sin embargo, el propio Searle (1983) no parece creer que sea posible la reducción completa de los verbos mentales a un lenguaje atómico de deseos/creencias (sería interesante analizar la posibilidad de descomposición semántica de los verbos mentales en rasgos de deseo, creencia y percepción, junto con otros de carácter más inherentemente connotativo, tales como los definidos tradicionalmente por el diferencial semántico de evaluación, potencia y actividad).

Es importante que destaquemos los siguientes puntos:

(a) Todos los lenguajes naturales contienen léxicos muy ricos y elaborados de verbos mentales, de tal manera que la presencia de verbos mentales puede considerarse un *universal lingüístico.*

(b) Existen sutiles diferencias entre los lenguajes en lo que se refiere a los rasgos que se codifican léxicamente o mediante sufijos en los verbos mentales. Así, por ejemplo, los Hopi contrastan, mediante sufijos verbales, los acontecimientos sensoriales objetivos y manifiestos, diferenciándolos de los de naturaleza subjetiva y no-sensorial. Los sucesos que el hablante ve o ha visto se formulan en forma «reportiva» (con independencia de que se produjeran en el pasado o se estén realizando en el presente). Los sucesos que el hablante piensa o cree que han sucedido se enuncian en un tiempo «expectativo», y las verdades generales en un tiempo «nómico» (Whorf, 1956; Paivio y Begg, 1981).

(c) Aunque estas diferencias, consideradas desde una perspectiva de relativismo lingüístico moderado, puedan facilitar, en mayor o menor grado, determinadas inferencias sobre lo mental, es probable que todos los lenguajes expresen un aparato conceptual y mecanismos computacionales complejos y universales, que asegurarían a todos los miembros de nuestra especie niveles básicos de pensamiento, conceptualización e inferencia acerca de su propia mente y la de los congéneres.

(d) Finalmente, una característica importante —y que ha sido destacada con frecuencia— de los verbos mentales es que tienen un carácter *transitivo.* Son intencionales, (se) refieren *a algo,* a un contenido, bien sea éste un objeto (Percibir (x)) o una proposición (Pensar que (y)). Expresan una *relación* entre un *sujeto* (generalmente un organismo) y un objeto que, con independencia de su referencia a un «mundo externo» es, en primer término, un *contenido mental* que se da en ese organismo.

Podríamos decir que la mayor parte de los verbos mentales «tienen siempre los brazos abiertos», en espera de un objeto al que abrazarse. Contienen un paréntesis vacío que tiene que ser llenado por una variable que especifica un estado *interno* y *representacional* del organismo (quizá haya verbos mentales no-intencionales, como pretende Searle, 1983. Pero éstos serían la excepción más que la regla, y parece que estaría restringido al campo semántico de las emociones). Pensar es siempre pensar en algo; percibir es percibir algo; no es posible recordar o desear si no es algo lo que se recuerda, algo lo que se desea.

Los verbos mentales, las actitudes proposicionales, implican relaciones de los organismos con el medio, pero éstas son relaciones mediatas. Relaciones que son, en primer lugar, de carácter *interno* y establecidas entre el organismo o sistema intencional y sus propios contenidos, sus propias representaciones.

Esa referencia a contenidos se produce tanto en los enunciados de creencia como en los de deseo. La diferenciación entre creencias y deseos no reside tanto en el hecho de que las primeras impliquen representaciones y los segundos no (pues, ¿cómo es posible la voluntad sin representación, a no ser en el mundo metafísico de Schopenhauer?; ¿cómo desear o querer a ciegas, sin algo que se desea o se quiere?), sino que más bien reside en el principio de que la «dirección de adaptación»

de las creencias va de las representaciones al mundo, y la de los deseos del mundo a las representaciones (Searle, 1983):

Si mis creencias resultan ser erróneas, son ellas las que fallan, no el mundo, como demuestra el hecho de que puedo corregir la situación simplemente cambiando mis creencias. Es responsabilidad de la creencia, por así decirlo, adaptarse al mundo, y cuando esa adaptación falla, la situación se resuelve modificando las creencias. Pero si no logro alcanzar mis intenciones, o mis deseos se quedan incumplidos, no puedo cambiar del mismo modo la situación modificando mis intenciones o deseos. En estos casos es, como si dijéramos, un fallo del mundo el no adaptarse a mis intenciones o deseos, y uno no puede cambiar las cosas diciendo que la intención o el deseo «estaban equivocados», en el mismo sentido en que podía cambiarlas reconociendo que las creencias eran erróneas. Las creencias, como las aseveraciones, pueden ser verdaderas o falsas, y podemos decir que tienen una dirección de adaptación «de la mente al mundo». Los deseos e intenciones, por otra parte, no pueden ser verdaderos o falsos, y podemos decir que tienen una dirección de adaptación que va «del mundo a la mente» (op. cit., p. 8).

Con arreglo a la distinción que establece Searle, podemos decir que el interés de los psicólogos cognitivos es, en esencia, el estudio de los actos mentales cuya dirección de adaptación va de la mente al mundo; en éstos, la naturaleza misma de las *representaciones internas* a través de las cuales se realiza tal adaptación se convierte en un problema fundamental, y adquiere aún más relevancia que en el caso de las actitudes proposicionales cuya dirección de adaptación es la inversa.

En uno y otro caso, el supuesto del que parte el vocabulario de las actitudes proposicionales, del que nos servimos en nuestras interacciones cotidianas con las personas, es el siguiente: «las personas tienen representaciones internas del mundo, y están en distintas clases de relaciones (i.e. de actitudes proposicionales) con respecto a tales representaciones».

Un segundo supuesto axiomático sería éste: «Las *razones* de las conductas intencionales de las personas (que no incluyen actividades tales como estornudar o producir reflejos parpebrales ante una bocanada de aire), es decir, *las razones de las conductas de las que cabe dar razones,* residen en formas de "relación con" y "manejo de" las representaciones».

En pocas palabras: el vocabulario de las actitudes proposicionales, que refleja nuestro modo de pensar acerca de nuestra propia conducta y la de los demás, es un vocabulario de naturaleza esencialmente *epistémica,* o, como dice Pylyshyn (1984), *cognitiva.* Presupone que, en la

forma elegante y sintética en que lo decía Descartes, *las personas obran por conocimiento* (y, en lo que tienen de objetos con mente, los animales). Hay un doble supuesto mentalista y «cognitivista» (en sentido laxo) omnipresente en la «psicofilia» de que hablaba Baroja, en esa «habilidad para hallarle su torcedor a cada uno» (Gracián, *Oráculo manual y arte de prudencia)* que forma parte de lo que algunos llaman «psicología natural» (Humphrey, 1983). La psicología natural no parece ser conductista, sino mentalista, y hace referencia constantemente a un vocabulario cognitivo.

6.2. La defensa cognitiva de las actitudes proposicionales

Recientemente, algunos de los defensores del paradigma C-R han reivindicado el valor, la importancia y el carácter irreductible del vocabulario mentalista, que forma parte esencial de la psicología natural de las actitudes proposicionales. Así, dos libros recientes que son ejemplares claros del paradigma, *Computation and Cognition* de Zenon Pylyshyn (1984) y *Psychosemantics* de Jerry Fodor (1988), dedican sus capítulos primeros a esa reivindicación. Debemos referirnos a ella, dada su significación para la comprensión del valor y la utilidad de la psicología cotidiana de las actitudes proposicionales, así como de la relación profunda que ésta tiene con la psicología cognitiva.

Pylyshyn (1984) se sirve de un ejemplo para ilustrar el valor y la significación del vocabulario cognitivo de la psicología natural: estamos en una calle y observamos una secuencia de acontecimientos. La secuencia consiste en que un peatón cruza la calle y, sin que lo advierta, un coche se le acerca velozmente. Luego se produce un frenazo brusco del coche y éste patina, dándose contra un poste. El peatón duda, se dirige al coche, mira por la ventanilla del conductor, echa a correr hacia una cabina telefónica y marca el 091.

¿Cómo puede describirse y explicarse esa secuencia de procesos?: «Parece —dice Pylyshyn— que una explicación adecuada de los sucesos en cuestión debería mencionar, entre otras cosas, que el peatón *percibió* el choque, que lo *reconoció* como un suceso que clasificó como accidente, que *dedujo* que podría haber heridos, que corrió a *averiguar* si de hecho había algún herido, que decidió pedir ayuda basándose en el *conocimiento* que tenía de cuál había de ser el tratamiento adecuado de las personas heridas, que *advirtió* la proximidad de una cabina de telé-

fonos, que *recordó* cuál era el número para casos de urgencia, y que marcó dicho número con la *intención* de pedir ayuda» (ed. 1988, p. 27). En suma, parece que una explicación adecuada requeriría el empleo de los verbos mentales y las actitudes proposicionales a los que nos hemos referido: percibir (x), reconocer (y), deducir (z), averiguar (h), decidir (i), advertir (j):

Aunque es posible que una explicación más técnica no utilizara estos términos concretos, o que utilizara algunos términos más o algunos menos, es evidente, en cualquier caso, que, para explicar el *porqué,* el *qué* y el *cómo* del suceso, parece que lo que realmente necesitamos es un vocabulario de este tipo, un vocabulario que no tiene nada que ver con el físico o el biológico (p. 28).

El vocabulario de los términos intencionales, que empleamos como psicólogos naturales para explicar la conducta, no puede *reducirse,* según esta perspectiva, ni al vocabulario físico, ni al biológico, ni a términos puramente conductuales. Una razón fundamental de ello es que el vocabulario cognitivo permite captar regularidades y generalizaciones que no podrían establecerse con los otros vocabularios (el físico, el biológico y el conductual).

Así, por ejemplo, cuando, al final de la secuencia anterior, el peatón asustado ha marcado el 0 y el 9, nuestros conceptos intencionales y cognitivos nos permiten realizar la predicción de que marcar el 1, puesto que *interpreta* la situación como una emergencia, *sabe* que el 091 es el número de la policía, *cree* que se puede pedir ayuda en situaciones de emergencia, etc. Pero esta predicción no podría realizarse mediante un vocabulario *estrictamente* físico, biológico o conductual.

Además, estos otros vocabularios incluirían en la misma categoría secuencias que serían diferenciadas por el vocabulario cognitivo: por ejemplo, una secuencia como la anterior, pero realizada como una escena de una película, no podría diferenciarse de la que sirve de ejemplo en términos de su descripción física, biológica y conductual; pero sería claramente diferente en su descripción intencional. En este caso, el peatón *fingiría* estar asustado, el conductor *simularía* un accidente, etc. (este ejemplo es una buena ilustración de la intuición, que desarrollaremos algo más, posteriormente, de que las capacidades de fingir, mentir y engañar determinan decisivamente la imposibilidad de explicar, de forma completa, la conducta en términos puramente extensionales o conductuales).

Por otra parte, el vocabulario puramente conductual categorizaría, como pertenecientes a clases diferentes, secuencias que podrían ser incluidas en una misma clase intencional: imaginemos, por ejemplo, que el peatón lleva el brazo escayolado y pide a otra persona que marque el 091. Necesitamos un vocabulario cognitivo para abstraer lo que hay de común en esta secuencia y la del ejemplo, del mismo modo que Tolman (1932; Lafuente, 1986) lo necesitó para explicar que sus ratas de laboratorio aprendieran a recorrer un laberinto *con independencia* de los movimientos concretos (de correr, nadar, etc) que las llevaban hasta una recompensa.

El hecho de que el vocabulario cognitivo permita establecer una categorización de una secuencia de acontecimientos que no se superpone a las categorías conductuales o neurofisiológicas (i.e. no se limita a reflejarlas, ni siquiera a «refinadas», sino que «traspasa» esas otras formas de categorizar) implica que permite también formular generalizaciones que no pueden seguirse ni de las descripciones de los movimientos globales del organismo ni de las que se realizan en términos de respuestas neurales o musculares específicas.

Ningún conjunto de descripciones en esos otros vocabularios, el de la fisiología y el de la conducta, puede corresponderse con algo tan simple como «el acto de marcar un uno», ni permite relacionar el conjunto de formas (potencialmente ilimitado) de aprender un número de teléfono, con el conjunto de formas (potencialmente ilimitado) de saber que se está ante una emergencia, y el conjunto de secuencias de movimientos (cuyo número es potencialmente ilimitado) que pueden corresponder a la acción de marcar el 091.

En definitiva, el único vocabulario que permite dar razón de la conducta de ciertos sistemas *en tanto que conocen* es el vocabulario cognitivo, con todo su aparato de actitudes proposicionales, procesos internos y supuestos acerca de las representaciones: «Las ciencias cognitivas —dice Pylyshyn— se apoyan en el supuesto fundacional de que existe un conjunto natural de generalizaciones que puede ser captado utilizando ésos términos cognitivos y, por tanto, que existe un dominio científico natural que constituye el objeto material de la ciencia cognitiva» (p. 35). La necesidad de ese dominio científico autónomo, que se define con un vocabulario cognitivo, depende, de forma esencial, del hecho de que ciertos organismos o sistemas complejos muestran formas de conducta que son notablemente independientes de las condiciones estimulares inmediatas o que, en la medida en que responden a éstas, lo hacen en

función de sus *significados* (y, por consiguiente, de ciertas estructuras *intencionales*) y no de sus propiedades físicas.

La independencia estimular de la conducta de algunos sistemas complejos (organismos o máquinas) no es, además, estocástica ni caprichosa, sino que está regida por principios diferentes que sólo emergen en las descripciones que se realizan con un vocabulario cognitivo: la conducta del que está triste y apartado en una fiesta, mientras los que le rodean cantan canciones regionales y beben champán, puede comprenderse, por ejemplo, cuando se *conoce* que el triste *desea* tener relaciones más estrechas con una de las bellezas de la fiesta, pero *infiere,* por lo que *percibe* en la complaciente conducta de ésta con un rival, que ella *prefiere* a este último, por lo que *piensa* que no existen *expectativas* de lograr sus *propósitos* en la fiesta. Las relaciones entre percepciones, creencias, deducciones, expectativas y propósitos configuran una red explicativa que permite dar cuenta de una conducta que resultaría extrañamente desajustada si no se incluyera en esa red intencional.

El grado de dependencia de la conducta de los sistemas con respecto a las propiedades físicas del entorno parece estar relacionado, a su vez, con la complejidad de tales sistemas como mecanismos de cómputo y representación. Por lo que sabemos, es mucho más fácil predecir la conducta de las ranas, partiendo de descripciones físicas, que predecir la de los chimpancés enanos o las personas. «Si intentamos describir la conducta humana en términos de las propiedades físicas del entorno, pronto concluiremos que, salvo en el caso de los desplazamientos, las caídas /../, la evitación de muros y determinados reflejos autónomos, la conducta humana es esencialmente aleatoria. Además sabemos que la conducta humana, si se describe en términos cognitivos, es muy regular y sistemática. Y éstas, naturalmente, no son buenas noticias para el conductismo», dice Pylyshyn (1988, p. 37).

Una manera diferente de formular la misma idea, que aparece en el párrafo anterior, consiste en decir que los organismos más complejos responden, de forma selectiva, a propiedades *no proyectables* del entorno, a propiedades que no admiten una caracterización física (como las propiedades de ser un uno, ser hermoso, ser una palabra o una oración en castellano, ser un zapato: ni siquiera los humildes zapatos pueden ser caracterizados como tales en términos escuetamente físicos).

Para Fodor (1984), la capacidad de responder a propiedades no proyectables constituye un criterio clave para distinguir entre los sistemas

que realizan inferencias y aquellos otros que se limitan a reaccionar causalmente a las propiedades del entorno.

La autonomía relativa de la conducta con respecto a la estimulación inmediata depende, en efecto, de la posesión, por parte del organismo o sistema que se conduce, de mecanismos de categorización e inferencia que le hacen capaz de ir «más allá de la información dada» (Bruner, 1974). Estos mecanismos permiten:

(1) Identificar (categorizar) los objetos, y sus propiedades y relaciones, como pertenecientes a las mismas clases de equivalencia.

(2) Hacer uso de redundancias del medio para predecir y «completar» sus paréntesis vacíos o borrados por el «ruido», en función de la categorización de los estímulos.

(3) Emplear reglas que especifican esas regularidades.

En suma, son mecanismos que implican la construcción de *códigos,* que no constituyen una proyección puntual de las características físicas del medio, y que contienen entidades (llámense palabras, conceptos, esquemas, categorías, relaciones abstractas, oraciones) que tienen la propiedad de ser *intencionales*, de ser —como dice Clark, 1989— *semánticamente transparentes,* de poseer un significado que se deriva no sólo de las propiedades físicas del medio, sino también de las interacciones complejas de sistemas también complejos con ese medio, y de las capacidades de *inferencia* que tales sistemas poseen (Bruner, 1974).

Con arreglo a lo que acabamos de decir, la pertinencia y necesidad de emplear un vocabulario cognitivo, para explicar la conducta de un sistema, depende de que esta conducta muestre una autonomía no-aleatoria con respecto a la estimulación inmediata del medio, y de que la relación mediata con éste pueda explicarse en términos de inferencias, que realiza el sistema, sobre representaciones establecidas en códigos no proyectables. Esta caracterización nos lleva al núcleo esencial que define lo que tiene en común la psicología cognitiva *natural,* de carácter netamente mentalista, que utilizamos como artefacto para saber a qué atenernos en nuestras interacciones con las personas y la psicología cognitiva *científica,* que trata de ser al tiempo, mentalista y objetiva, y que desarrollamos los psicólogos como profesionales de la mente.

Como ya señalaba en otra ocasión (Rivière, 1987), probablemente lo más general y común que podemos decir de la Psicología Cognitiva es que refiere la explicación de la conducta a entidades mentales, de naturaleza epistémica, para las que reclama un nivel de *discurso propio,* diferente de aquel que se limita a establecer relaciones entre eventos y

conductas externas (como en el vocabulario del análisis experimental de la conducta) y del referido a los procesos fisiológicos que subyacen a las funciones mentales: un nivel de discurso que contiene palabras tales como «esquema» y «operación» (Piaget, 1971), cómputo y proposición (Pylyshyn, 1989), prototipo y categoría (Rosch, 1978), imagen y representación analógica (Kosslyn, 1980).

El nivel de discurso propio, y diferenciado, que puede realizarse con este vocabulario depende, a su vez, del hecho de que el plano cognitivo posee una *autonomía funcional.* Es decir, en la conducta se dan regularidades y formas de determinación que van más allá de lo que puede expresarse mediante cadenas asociativas de izquierda a derecha, y que no pueden describirse, sin perder información relevante, en términos de relaciones probabilísticas conducta-medio, procesos neurofisiológicos, e influencias sociales o culturales.

Además, aunque esas regularidades puedan corresponder, de forma más o menos «realista», a regularidades que se dan en el mundo «real» (Gibson, 1966; Neisser, 1976), no pueden expresarse por relaciones *puntuales* entre variaciones de energías físicas y variaciones de la conducta del organismo o sistema. Ello es así en la medida en que los sistemas complejos se hacen capaces de *desplazarse* espacial y temporalmente del medio inmediato, manejando representaciones, y también en la medida en que emplean mecanismos «Top-down», que determinan el procesamiento de los estímulos del medio en función de los conocimientos codificados en los sistemas de memoria. En este sentido, podemos decir que los sistemas cognitivos requieren, necesariamente, mecanismos de *memoria* y *predicción,* y sistemas de inferencia capaces de abstraer las propiedades invariantes del medio, que subyacen a las transformaciones más puntuales y versátiles de las energías físicas que transducen sus receptores sensoriales (Rivière, 1987).

Los supuestos que establecen que, para dar cuenta de la conducta de sistemas adaptativos complejos, es necesario atribuirles capacidades de representación, inferencia, memoria y predicción, son comunes a la psicología natural y la psicología cognitiva. Sin embargo, las semejanzas de vocabulario y de algunos supuestos no deben ocultarnos las diferencias importantes que existen entre una y otra: la psicología natural no es, ni trata de ser, mecanicista. Sus conceptos son artefactos predictivos de la conducta de las personas en especial, y no permiten formular directamente *leyes nomológicas,* sino inferencias idiográficas. Además, las categorías y formas de explicación de la psicología natural están completamente apegadas a la pregnante fenomenología de la conciencia, y dependen —como comentaremos más extensamente en otro momento— de una suposición fundamental de *identidad intersubjetiva* entre las personas, que es ajena a las posibilidades de explicación *objetiva* de la conducta, que pretende alcanzar la psicología cognitiva.

Sin embargo, las categorías de la psicología natural sirven de horma primera para la definición del vocabulario cognitivo y, como señala Pylyshyn (1984), «nuestras nociones preteóricas acerca de qué sea lo que constituye el dominio de la psicología cognitiva han sido configuradas por nuestra experiencia en el tratamiento de ciertos tipos de regularidades empíricas, de forma que las nociones mantienen cierta validez primaria, al menos en una primera aproximación. De esta forma, consideramos como psicológicas unas regularidades que, por ejemplo, relacionan propiedades «percibidas» /.../ con acciones intencionales y dotadas de significado /.../, en tanto que consideramos como no psicológica, al menos en parte, la relación que existe entre tropezar en una piedra y golpearse la rodilla» (ed. española de 1988, p. 43).

Así, los mecanismos de comprensión y predicción de la conducta, de que se sirve la psicología natural mentalista (que no es lo mismo que la «psicología popular», como aclararemos en su momento), y que se articulan alrededor del lenguaje y el sistema conceptual de las «actitudes proposicionales», han jugado siempre un papel decisivo en el desarrollo de la psicología científica: Primero, como determinantes de la propia pretensión de hacer una psicología, como ciencia autónoma, y no reductible a la física o la biología; segundo, y paradójicamente, como *obstáculos epistemológicos* para la elaboración de una psicología científica (posteriormente aclararemos este punto) y, en tercer lugar, como mecanismos correctores y críticos que siempre han intervenido, más o menos conscientemente, en los procesos de redefinición y re-dirección de la psicología científica, debido a que las categorías de la psicología natural están profundamente interiorizadas y definen un paisaje conceptual que, querámoslo o no, sirve de fondo a los esfuerzos por desarrollar una ciencia objetiva de la conducta o de la mente.

Pylyshyn (1988) destaca varios aspectos en que la psicología natural, mentalista e informalmente «cognitiva», condiciona la elaboración de la psicología cognitiva propiamente dicha:

(1) Establece una taxonomía de funciones mentales que delimita, en un primer momento, los campos de investigación (los libros de psicología cognitiva se dividen en capítulos que corresponden, *grosso modo,* a as-

pectos de la taxonomía de los verbos mentales y las actitudes proposicionales: memoria, comprensión, solución de problemas, razonamiento, etc.).

(2) Define formas de explicación, en función de razones, conocimientos, etc, que implican el uso de un vocabulario cognitivo.

(3) Este vocabulario produce una psicología natural razonablemente acertada y predictiva en situaciones concretas.

(4) La psicología natural se corresponde con la intuición de racionalidad en que se basan las interacciones humanas a las que cabe imputar responsabilidad, y en que se basan también las ciencias cognitivas.

La relación entre la psicología cognitiva científica y la psicología natural de las actitudes proposicionales ha sido, como es fácil entrever por lo que llevamos dicho, una relación ambigua: desde una perspectiva histórica, el enfoque cognitivo supuso la recuperación del vocabulario mentalista y epistémico de la psicología natural, así como de algunas de sus intuiciones fundamentales, como las relacionadas con la racionalidad (no necesariamente lógica) y dependencia cognitiva de gran parte de la conducta humana.

Por otra parte, esa recuperación apenas ocultaba la presencia determinante, en la psicología cognitiva, de axiomas opuestos a las intuiciones de la psicología natural: para ésta no resulta intuitivamente obvia la naturaleza inconsciente de la mayor parte de la maquinaria computacional de la mente, ni siquiera la existencia de tal maquinaria *computacional*,, es decir, semejante —literal o metafóricamente— a la de los sistemas artificiales de cómputo.

En oposición a los psicólogos y filósofos que postulan que las categorías de la psicología natural definen concepciones de lo mental vacías o, peor aún, fundamentalmente erróneas (Stich, 1983; Churchland, 1979), Fodor (1988) llega a afirmar lo siguiente:

> si la psicología intencional de sentido común llegase, realmente, a colapsarse, estaríamos, sin comparación, ante ia mayor catástrofe intelectual de la historia de nuestra especie; si estuviéramos equivocados acerca de la mente, ésta sería la mayor equivocación que habríamos tenido nunca acerca de algo. E! colapso de lo sobrenatural, por ejemplo, no podría compararse con éste; el teísmo nunca ha estado tan íntimamente implicado en nuestro pensamiento y nuestras prácticas —especialmente en nuestras prácticas— como lo están las explicaciones en términos de creencias/deseos. Nada, con excepción quizá de nuestra física de sentido común —nuestro compromiso intuitivo con un mundo de objetos de tamaño intermedio, independientes del observador—, está tan cerca de nuestro núcleo cognitivo como lo están las explicaciones intencionales (p. 12).

Para Fodor (1988), hay tres compromisos fundamentales que presupone la psicología natural y que se conservan en el enfoque cognitivo de la mente:

(1) En primer lugar, las actitudes proposicionales tienen la propiedad de ser *semánticamente evaluables.* Es decir, mantienen una relación con el mundo tal que permite determinar cuando se cumplen o no (cuando son verdaderas o falsas las creencias, satisfechos o insatisfechos los deseos, ciertas o erróneas las intuiciones, etc). Las actitudes proposicionales son entidades cuyos contenidos se satisfacen o no por estados de cosas en el mundo, siendo el modo de satisfacción dependiente de la forma de relación que, por definición, establecen entre el sistema que las tiene y el mundo con el que ese sistema se relaciona (creer es un modo de relación diferente de desear, y éste lo es de temer, etc).

(2) Además, «es característico de la psicología del sentido común de los deseos y creencias /.../ el hecho de que atribuye contenidos y poderes causales *precisamente a las mismas cosas que toma como semánticamente evaluables»* (Fodor, 1988, p. 12). Esto significa, en otras palabras, que la psicología natural establece inferencias causales (quizá sería mejor decir *racionales)* por medio de un aparato conceptual que, para realizar dichas inferencias, se sirve muy fundamentalmente de dos tipos de elementos: los contenidos de las representaciones, por una parte, y las formas de relación del organismo con dichas representaciones (si yo creo que «x sospecha que (a), x no desea (a), x cree que y es responsable de (a), y creo también que y no es responsable de (a)», entonces inferiré que x imputará equivocadamente (a) a y, y tendrá quizá una actitud negativa hacia y).

(3) En su aspecto más global, el tipo de generalizaciones racionales que establece la psicología natural se preservan —en cuanto a su forma— en la psicología cognitiva: «Una psicología explícita que reivindique las explicaciones en términos de deseos y creencias del sentido común —dice Fodor— debe permitir la asignación de contenidos a estados mentales causalmente eficaces, y debe reconocer explicaciones comportamentales en que las generalizaciones se refieren a (o se cuantifican sobre) los contenidos de los estados mentales que incluyen /.../. La psicología del sentido común no será reivindicada a menos que resulte ser, cuando menos, aproximadamente cierta» (1984, p. 15).

Con independencia de que la psicología cognitiva implique (como pretende Fodor) o no una reivindicación «científica» de los objetos y relaciones mentales que se emplean en la psicología natural de las acti-

UNIVERSIDAD DE ANTIOQUIA
BIBLIOTECA CENTRAL

tudes proposicionales, lo que sí es cierto es que la comprensión de la significación de esos objetos y relaciones, i.e. de las funciones mentales tal como éstas se comprenden en las interacciones cotidianas, es un recurso heurístico muy poderoso para comprender los orígenes y fundamentos de nuestra ciencia. Por ello, parece conveniente que resumamos, antes de seguir adelante, las conclusiones que se siguen de las reflexiones recogidas en este capítulo. Son las siguientes:

(1) Hay ciertos sistemas que parecen «guiarse por conocimientos» en su forma de comportarse. Son sistemas complejos, cuya conducta no parece ser una proyección literal de las variaciones puntuales de las energías del medio. Sistemas que van más allá de la información dada, gracias a que cuentan con mecanismos para conservar información, abstraer aspectos invariantes del medio y predecir variaciones y regularidades de éste.

Por muy elementales que sean, todas esas capacidades implican el empleo de procesos de *codificación*, que construyen clases de equivalencia de los estímulos ambientales. Además, de esos sistemas podemos decir que realizan *inferencias*, manipulando las relaciones entre objetos representacionales que están descritos en el lenguaje de los códigos que poseen. Sólo puede decirse de un sistema que se guía por conocimientos cuando su conducta evidencia la influencia de mecanismos «top-down», de arriba a abajo, que van de lo codificado previamente a lo que se está codificando, del «interior» del sistema a la «interpretación» de las aferencias que éste recibe, de las formas de organización del sistema a sus formas de conducta y recepción de los estímulos.

(2) Un criterio fundamental para determinar si un sistema se guía por conocimientos es el grado de independencia no-aleatoria que muestra dicho sistema con respecto a los estímulos inmediatos del medio, descritos en términos físicos. Los sistemas cognitivos responden a propiedades no-proyectables de los estímulos y alcanzan un grado de *autonomía funcional* que no poseen los sistemas no-cognitivos. Una función esencial que realizan los sistemas cognitivos consiste en abstraer los aspectos invariantes del medio que resultan relevantes desde el punto de vista de sus necesidades adaptativas.

Los mecanismos de abstracción de invariantes son, en primer término, perceptivos (como en la definición de las constancias perceptivas), pero desbordan enseguida la percepción —en el hombre y otros organismos superiores— para configurar, en primer lugar, un mundo de objetos permanentes y, luego, de relaciones estables, más allá de las apariencias perceptivas (Piaget, 1961).

(3) Hay organismos que, además de *tener* esas propiedades cognitivas, son capaces de *atribuir* a otros organismos su posesión. Hay datos que indican que los antropoides superiores son capaces de realizar esas atribuciones (Premack y Woodruff, 1978), pero, sin duda, esa capacidad se elabora y refina enormemente en el hombre. Sirviéndose de los mismos *principios* cognitivos (aunque no necesariamente de los mismos mecanismos) que permiten abstraer invariantes del medio, inferir en él relaciones, codificar sus objetos formando clases de equivalencia, el hombre abstrae invariantes de su propia conducta y la de sus congéneres, infiere relaciones internas que guían esa conducta, y codifica de un modo especial la conducta de los demás.

Así, los miembros de nuestra especie no sólo son sistemas cognitivos sino que presuponen que los otros lo son. No sólo se sirven de su mente cognitiva para adaptarse mejor al medio físico, sino también —y muy significativamente— para adaptarse al medio social.

(4) El sistema conceptual y lingüístico que subyace a esa adaptación tiene características especiales. Se basa en un vocabulario peculiar, mentalista y cognitivo, en que los verbos mentales y las actitudes proposicionales ocupan un lugar decisivo. Estos artefactos conceptuales consisten en descripciones de formas *internas* de relación de sistemas complejos (generalmente personas) con representaciones. Poseen así una propiedad esencial de *intencionalidad.*

Las entidades a las que se refieren los elementos del vocabulario mentalista de la psicología natural se entienden como «semánticamente transparentes», o «semánticamente evaluables», y se encadenan en relaciones de causa-efecto de carácter teleológico y, frecuentemente, propositivo. Hay razones importantes para suponer que los elementos conceptuales de la psicología de las actitudes proposicionales, que son enormemente complejos y diversos, cumplen la función esencial de servir de soportes para la *prediccion y comprensión* de la conducta de los miembros de la propia especie, y para la atribución a éstos de estados internos.

En cualquier caso, los sistemas de inferencia «meta-representacional», así constituidos, se articulan en torno a un axioma básico: que las personas actúan, la mayor parte de las veces, en función de: (a) los contenidos de sus representaciones, (b) sus formas de relación con esos contenidos (llamando formas de relación a cosas tales como «creer», «intuir», «recordar», «sospechar», «desear», etc), y (c) las relaciones, de segundo orden, entre tales formas de relación.

(5) Finalmente, el desarrollo interno de la psicología cognitiva y, más

en general, de la psicología científica sólo puede entenderse desde el marco de esa psicología natural que, al tiempo que parece reivindicar la existencia de un plano psicológico autónomo y ofrecer instrumentos conceptuales para un primer acercamiento a él, contiene un vocabulario y unos supuestos implícitos que parecen ofrecer obstáculos importantes para su constitución: un vocabulario ligado a la subjetividad, de naturaleza no-extensional, que permite dar razones más que causas; unos supuestos de teleología en la acción, reducción «homuncular» de muchas de las explicaciones y libre indeterminación de la conducta.

Todo ello explica la ambigua relación que existe entre la psicología cognitiva y la psicología natural del sentido común: aquella se sirve frecuentemente del vocabulario de ésta, pero introduce sutiles modificaciones, tan peligrosas para la comprensión mutua como las que existen entre palabras semejantes fonéticamente —pero de significado matizadamente diferente— entre el español y el italiano, por ejemplo. Además, frecuentemente la psicología cognitiva usa ese vocabulario para cuestionar axiomas básicos en la psicología de sentido común: axiomas como el de la eficacia causal de la conciencia," la transparencia a ella de procesos mentales, y la racionalidad lógica de buena parte de los procesos de inferencia. En su ambigua relación con la psicología natural, la psicología cognitiva, al mismo tiempo, la reivindica y la niega, conserva su vocabulario y modifica sutilmente su significado, al introducirle en el engranaje mecanicista de la mente de la que habla.

Capítulo 7

LA ATRIBUCION DE LO MENTAL. CONSIDERACIONES SOBRE LA PSICOLOGIA NATURAL

7.1. El lenguaje de lo mental: un diálogo con Archer

A pesar de que los psicólogos seamos profesionales de las funciones mentales, y tengamos una vieja querencia con relación a ellas, hasta muy recientemente no hemos caído en la cuenta de la importancia psicológica que tiene el hecho de que *los humanos no sólo tenemos funciones mentales sino que nos las atribuimos a nosotros mismos y se las atribuimos a los demás.* Una tesis fundamental que me gustará reflejar en estas páginas es la idea de que gran parte de las funciones mentales humanas, y muy en especial las de pensamiento y lenguaje, sólo pueden explicarse teniendo en cuenta esa capacidad humana de realizar funciones mentales acerca de las funciones mentales propias y ajenas, de tener, por así decirlo, funciones mentales de segundo orden.

Probablemente, esa capacidad ha pasado inadvertida durante tanto tiempo por una razón paradójica, a saber, que en psicología lo obvio se hace fácilmente *invisible.* Sucede con algunas habilidades y destrezas psicológicas algo semejante a lo que sucede con algunas personas que, ocupando silenciosa y permanentemente un cierto lugar, llegan a pasar inadvertidas y a «desapercibirse», como si formasen parte del mobiliario. De forma parecida, los estudios sobre la percepción y el lenguaje han tenido que enfrentarse con el obstáculo que ha supuesto el hecho de que funciones muy complejas, tales como las de percepción visual de objetos tridimensionales y comprensión de oraciones gramaticales de un lenguaje, suelen producirse de forma tan «silenciosa», eficaz, y aparentemente fácil que resultan casi invisibles.

Una de las razones de la importancia de la obra de Chomsky. (1957, 1965) en el estudio del lenguaje, y de la de Marr (1982) en el de la visión, reside en que estas obras nos han enseñado una estrategia importante a los psicólogos cognitivos: la estrategia que consiste en *buscar lo computacionalmente complejo por debajo de lo fenomenológicamente obvio.* El psicólogo cognitivo ha aprendido a desconfiar, en el mejor sentido de la palabra, de la suavidad y aparente falta de esfuerzo con que se realizan funciones mentales de enorme complejidad, tales como comprender el discurso, ver objetos de tres dimensiones, producir oraciones gramaticales, recordar los significados de lo que nos dicen, etc.

Parece que ha llegado la hora de «desconfiar» también de la pasmosa facilidad con que las personas realizamos cotidianamente actividades tales como atribuir intenciones a otros, inferir lo que piensan, creen y desean, comunicarles nuestros pensamientos y comprender los de los demás, y adaptarnos a lo que saben e ignoran. Todavía sabemos muy poco de los mecanismos computacionales que subyacen a estas destrezas, cuya pregnancia fenoménica es tan grande que, paradójicamente, ha llegado a ocultarlas como fenómenos dignos de explicación científica. Son destrezas que ponemos constantemente en juego en nuestra vida de relación con los demás, y que sirven de fundamento a la competencia comunicativa.

Cualquier situación comunicativa podría servir para ilustrar esa fácil complejidad con que se desarrollan conceptos e inferencias acerca de la mente y el conocimiento, en nuestras interacciones habituales. Un buen ejemplo, elegido por un tanteo muy breve y azaroso, puede ser el que se presenta en la conversación que veremos a continuación, tomada de la primera novela a mano, *El otro lado del dólar* de Ross MacDonald: un detective inolvidable, Lew Archer, se encuentra en una situación en que tiene que poner en juego sus dotes de «psicólogo natural» o, si se quiere, de «psicofílico». La situación consiste en que un joven, Tom Hillman, se ha fugado de una residencia-escuela psiquiátrica para delincuentes juveniles, y Lew Archer, contratado para localizarle, informa a los asustados padres de Hillman de ciertos hechos, en el siguiente tenso diálogo:

Le dije a Hillman:

—¿Puedo hablarle a solas? He descubierto algo que debe saber.

—Puede decírmelo delante de Elaine, y también de Dick.

Advertí que Leandro estaba al lado de la puerta.

• —Prefiero no hacerlo.

—Pero no es el que decide —era un curioso eco del hombre del teléfono—. A ver lo que averiguó.

—A su hijo lo vieron acompañado de una mujer casada, apellidada Brown. Es rubia, tipo de actriz o bailarina, mucho mayor que él y parece que le sacaba o quería sacarle el dinero /.../

Elaine levantó las manos abiertas frente a su cara, como si tantas palabras le confundieran.

—¿Qué es eso de «acompañado»?

—Salía con esa mujer. Los vieron juntos ayer por la tarde /.../

Elaine Hillman alzó la mirada. Parecía horrorizada.

—¿Insinúa que Tom tenía algo con ella?

—Estoy informando sobre los hechos.

—No creo ni en uno solo de sus hechos.

—¿Cree que estoy mintiéndole?

—No adrede, quizá. Pero debe haber algún terrible error.

—Opino lo mismo —dijo Dick Leandro desde el umbral—. Tom siempre fue un muchacho de vida intachable.

Hillman no dijo nada. Tal vez supiera algo de su hijo que los otros ignoraban.

(Ross MacDonald, *El otro lado del dólar,* 1980, p. 83).

Los finales de las desventuras de Tom Hillman, los desvelos de sus padres, y los descubrimientos de Archer, deben consultarse, desde luego, en la fuente original. No son éstos nuestros desvelos de ahora. Lo que nos importa destacar es la elaborada red de verbos intencionales y cognitivos que se emplean en la conversación anterior: *decir, descubrir, decidir, averiguar, ver, insinuar, informar, creer, mentir, opinar, saber* e *ignorar.* Además están los verbos preferir y querer, que también tienen una naturaleza intencional, pero no tan netamente epistémica, y otro verbo muy significativo, hablar, que no tiene carácter mental, ni intencional, ni epistémico (ello explica el persistente intento objetivista de reducir «pensar» y «decir» a «hablar»).

Por otra parte, dentro de los verbos intencionales y epistémicos, podemos hacer una distinción reveladora entre aquellos que son de naturaleza *interna* y *mental,* y los que son de carácter esencialmente externo. Los primeros se relacionan con el pensamiento y la percepción: «descubrir», «decidir», «averiguar», «ver», «creer», «saber» e «ignorar». Los segundos con el lenguaje y sus funciones: «decir», «insinuar», «informar», «mentir», «opinar». Existe una delicada relación entre los primeros verbos y los segundos. Searle (1983) ha expresado esta relación se-

ñalando que los *estados intencionales* (a los que hacen referencia los primeros verbos) *definen las condiciones de sinceridad de los actos de habla* (a los que se refieren los segundos verbos): «en la realización de cualquier acto ilocucionario que tenga un contenido proposicional —dice— expresamos un cierto estado intencional con ese contenido proposicional, y el estado intencional constituye la condición de sinceridad de ese tipo de acto de habla. Así, por ejemplo, si yo hago la aseveración de que p, expreso la creencia de que p. Si hago la promesa de hacer A, expreso la intención de hacer A» (p.9).

La idea de que los estados mentales intencionales se corresponden con las condiciones de sinceridad de los actos ilocucionarios es una expresión precisa de un concepto más general, a saber, que el lenguaje es, antes que ninguna otra cosa, *un sistema para explicitar y negociar estados mentales*, un sistema que permite la comunicación entre las mentes (una evidencia que está sirviendo de guía para los desarrollos recientes de una psicología del lenguaje de fuerte impregnación pragmática y, a la vez, cognitiva, como la que proponen, en *Relevance*, Sperber y Wilson, 1986).

Si la función más básica y profunda del lenguaje es la explicitación y negociación de estados mentales intencionales, es lógico que todos los lenguajes cuenten con ricos depósitos de verbos mentales y artefactos con los que hacer manifiestas las actitudes proposicionales. Podemos decir que el lenguaje tiene la virtualidad de hacer salir a la mente de su reducto intrasubjetivo, exponiéndola a una consideración intersubjetiva.

La mente es, así, algo que *se dice* de ciertos objetos, y todo nuestro lenguaje está impregnado de ella. Sería, el lenguaje, una jerga incomprensible para un ser que no partiera de esa impregnación, de ese supuesto de que es una mente que se comunica con otras mentes. Para alguien que no fuera capaz de presuponer —aun sin percatarse de ello— que, como él mismo, sus interlocutores son seres que descubren, deciden, averiguan, ven, insinúan, creen, mienten, opinan, saben e ignoran. Hemos dicho: «Presuponer que, *como él mismo*, sus interlocutores...», y eso equivale a decir que *el uso y aprendizaje del lenguaje presupone la noción esencial de que los otros son seres con mente intencional, cuya estructura esencial (i.e. las capacidades de creer, pensar, averiguar, etc.) es idéntica a la de la mente propia.*

El uso lingüístico entre humanos parte, así, de la presuposición esencial de la identidad fundamental entre las mentes.

La observación anterior nos conduce a algunas consideraciones, algo más especulativas que las que hemos hecho hasta ahora, pero que merecen unas líneas antes de recuperar el hilo principal de nuestro argumento. La primera es que, cuando hablamos de presuposición, lo hacemos también en un sentido genético; es decir: no sería posible la adquisición de un lenguaje sin una cierta capacidad previa de penetración intersubjetiva, en la mente de otros. Es cierto que el lenguaje mismo va a ser luego un instrumento muy poderoso de elaboración, transformación y refinamiento de esa capacidad intersubjetiva, pero su desarrollo exige, cuando menos, ciertas destrezas previas de contagiarse de los estados emocionales de otros (lo que Colwin Trevarthen llama «intersubjetividad primaria») y poseer una noción, cuya importancia genética no es, en absoluto, menor que la que tiene la noción de «objeto permanente»: la noción de identidad esencial entre las mentes (relacionada con el concepto de «intersubjetividad secundaria» de Trevarthen; vid. 1977, 1982).

La segunda consideración es que ninguna formulación estrictamente conductista y objetivista puede dar explicación al hecho de que los humanos seamos capaces de emplear un *lenguaje privado* (empleando aquí el término en el sentido de un lenguaje que no posee referentes externos, tal como el que se utiliza cuando se usan los verbos «pensar», «averiguar», «imaginar», «sospechar», «creer») y ninguna explicación conductista de la adquisición del lenguaje puede, si se mantiene en sus términos explicativos propios, dar cuenta de cómo es posible que se *aprenda* un vocabulario como el de los verbos mentales. La dificultad para justificar la existencia y adquisición del lenguaje privado, desde una perspectiva estrictamente conductista, ya fue vista por Wittgenstein (1968) y ha dado lugar a un interesante debate filosófico (vid. la revisión editada por Villanueva, 1979). Aquí sólo nos interesa destacar que las formulaciones de la adquisición del lenguaje como proceso de asignación (por asociación condicionada, «juego de lenguaje» o el mecanismo que sea) de términos a referentes *externos, intersubjetivos* y *objetivamente compartidos,* tiene que enfrentarse al hecho del lenguaje mental privado como una *anomalía explicativa* muy difícil, o imposible, de asimilar en el aparato conceptual conductista.

La idea que estamos formulando aquí está muy lejos de ese aparato conceptual: si pueden definirse evolutivamente mecanismos de penetración intersubjetiva, previos al lenguaje (tales como los expresados por la notable capacidad de compartir emociones e imitar de los niños de pocos meses), resulta fácil explicar que los niños puedan adquirir un vocabulario mentalista, que remite a procesos de naturaleza interna. Lo que ningún modelo conductista puede negar es el *fenómeno* comprobado

de que, desde el segundo año de vida, los niños desarrollan un vocabulario mental muy rico (Bretherton y Beeghly, 1982) y lo van refinando, mediante distinciones semánticas sutiles, como la que existe entre «saber» y «adivinar» (Miscione, Marvin, O'Brien y Greenberg, 1978) a lo largo del periodo básico de adquisición del lenguaje.

Ese vocabulario mental incluye tanto verbos epistémicos (Olson y Torrance, 1987; Wellman y Johnson, 1979; Johnson y Wellman, 1980; Wellman y Estes, 1987) como verbos emocionales (Bretherton, Fritz, Zahn-Waxler y Ridgeway, 1986), y se emplea en la realización de inferencias cada vez más complejas acerca de los estados mentales de otros.

Si bien para la adquisición del lenguaje, en general, es preciso el requisito evolutivo de la noción de que los otros son seres con una mente de estructura fundamental idéntica a la propia —y ello implica algún grado y modo de acceso a la propia mente—, para el desarrollo y refinamiento del lenguaje mental se requiere otra noción mucho más compleja, a saber: la de que los *objetos* o *contenidos* de las actitudes proposicionales de otros pueden ser *diferentes* de los que tienen las actitudes proposicionales propias, lo cual implica aprender a distinguir los estados mentales de otros de los propios. Es decir, si representamos las funciones mentales en forma de F (p), en que F es una relación con un contenido y (p) el contenido mismo, los usos lingüísticos y conceptuales de la psicología natural se basan en dos axiomas esenciales:

(1) La estructura de F es esencialmente idéntica en mí y en aquel con el que me relaciono o comunico.

(2) El contenido de (p) no tiene por qué coincidir en mí y en aquel con el que me relaciono o comunico.

De hecho, muchas actividades humanas de pensamiento y lenguaje contienen estrategias cuya función esencial es reconstruir estados mentales de otros, diferenciándolos de los propios, i.e. reconstruir representaciones acerca de otros que tienen la forma F (p) —en psicología evolutiva se emplea el término equívoco «teoría de la mente», para definir la capacidad de realizar tales reconstrucciones—. Para ilustrar este tipo de estrategias, puede servirnos el ejemplo del texto de Ross MacDonald que citábamos algunos párrafos más arriba.

Detengámonos, por ejemplo, en la última observación que se hace en el texto. Es la más misteriosa de todas: «Hillman no dijo nada. Tal vez supiera algo de su hijo que los otros ignoraban». Aquí se entremezclan varios verbos mentales en el proceso de pensamiento que MacDonald atribuye a Archer. Y se entrelazan de forma muy compleja e inferencial. Lo que se implica es esto: «(*Puesto que)* Hillman *no dijo* nada, (Archer *sospechaba* que) tal vez (el señor Hillman) *supiera* algo de su hijo que los otros (la señora Hillman y Dick Leandro) *ignoraban».* En suma, se trata de una inferencia compleja, por medio de la cual: (1) se atribuye un estado mental a un personaje A (*sospechar*) que implica, a su vez, una inferencia sobre (2) el estado mental de otro personaje B (*saber* algo), basando esa inferencia en (3) la *no* emisión de símbolos intencionales (*no decir nada),* por parte de B, y (4) haciendo consistir la inferencia de A en una diferenciación entre el estado mental que se atribuye a B (*saber* algo) y el que (5) se atribuye, en segundo grado, a C y D (*ignorar* algo). Es evidente que toda esta cadena de inferencias se basa en la presuposición o axioma previo de que las mentes pueden distinguir sus respectivos estados mentales.

«¡Todo esto es muy complicado! —podría decirse— ¡Pobre del lector de Ross MacDonald si tuviera que dedicarse a *inferir* que MacDonald *atribuye* a Archer la inferencia de que quizá Hillman *supiera* algo que los otros *ignoraban!»* Pues bien: eso es precisamente lo que hace, sin ningún esfuerzo aparente, y sin darse mucha cuenta, el lector de MacDonald. Y lo hace, además, una y otra vez. Por ejemplo, en la conversación que estamos comentando, el lector se enfrenta a un intercambio como este:

—¿Insinúa que Tom tenía algo con ella?
—Estoy informando sobre los hechos.
—No creo ni en uno solo de sus hechos.
—¿Cree que estoy mintiéndole?
—No adrede, quizá. Pero debe haber algún terrible error.

La posibilidad de comprender esta interacción verbal se basa en una delicada diferenciación semántica entre verbos cognitivos y términos epistémicos: no es lo mismo *insinuar* que *informar. No creer* no tiene por qué ser necesariamente un resultado de que el otro *mienta* (es decir, haga creer lo que él mismo no cree), sino que puede serlo de un terrible *error* (i.e. de tener una creencia cuyo contenido intencional no corresponde a la realidad del objeto al que se refiere). De nuevo se hace necesaria la intuición de la complejidad por debajo de la facilidad aparente con que el lector comprende esa interacción verbal entre Lew Archer y la señora Hillman. Esa comprensión se basa en competencias semánticas que permiten distinguir rasgos intencionales de matices muy

tenues, pero que tienen una enorme importancia en el plano de la mente que sirve de guía para orientar las interacciones cotidianas.

7.2. Los primates mentalistas y las ventajas evolutivas del engaño y de su reconocimiento

¿Por qué desarrollan los humanos un plano tan complejo de la mente como base de su «psicología natural»? Nicolás Humphrey (1983) ha especulado con la idea de que las condiciones del ambiente *social*, en la evolución humana, fueron los determinantes básicos del desarrollo filogenético de las competencias naturales de cómputo que definen la inteligencia humana: «Fueron las circunstancias de la vida *social* del hombre primitivo —el pertenecer a una comunidad humana con interacciones complejas su necesidad de ayudarse mientras al mismo tiempo ayuda a los demás— las que, más que nada, hicieron al hombre como especie, la criatura astuta y penetrante que hoy conocemos» (p. 14, ed. esp. 1987).

Las condiciones sociales de la evolución humana favorecieron el desarrollo de un «homo psychologicus», dice Humphrey, calculador y especialmente dotado en sus capacidades de inferencia interpersonal, de las que derivarían (y no al contrario) las competencias de inferencia impersonal. Esas capacidades de inferencia y cálculo del «homo psychologicus» se basarían, a su vez, en la estrategia de *emplear un cuadro privilegiado del propio yo como modelo de lo que es ser otra persona.*

Esta observación (que se corresponde con la que hacíamos antes de que la psicología natural se basa en un supuesto de identidad esencial entre las mentes) es importante porque nos permite intuir hasta qué punto existe una estrecha relación entre el desarrollo y la función de la conciencia reflexiva humana y el de las capacidades de inferencia y cómputo sobre el mundo mental de los demás. Si la suposición de Humphrey es correcta (y hay razones importantes para suponerlo), a lo largo de la evolución humana, se favorecieron las posibilidades de preservación de los rasgos genéticos de los individuos más capaces de acceder a su propia mente, y, en-consecuencia, de comprender y predecir la conducta de sus congéneres mediante inferencias mentalistas:

Para el psicólogo natural —dice Humphrey—, la autoobservación significa no sólo observar desde el exterior sino observar desde dentro, no sólo contemplar él propio comportamiento, sino contemplarlo en los «pensamientos y las pasiones» que lo acompañan. Y la capacidad para este tipo de observación interna —la capacidad de mirar en sí mismo y considerar lo que hacemos al pensar, tener esperanzas, temores, etc.—/.../ a mi parecer, representa el desarrollo más peculiar y refinado de la mente humana. /.../ Para darle un nombre, digamos que es la capacidad de «conciencia reflexiva»: una conciencia de la conciencia (p. 16).

La perspectiva de Humphrey se basa en una descripción, de los factores decisivos en la evolución humana, bastante diferente de la que ha predominado en las explicaciones tradicionales. Según esta nueva descripción (muy afín a las que vienen desarrollando, en los últimos años, otros investigadores, desde la paleontología, la etología y la antropología evolucionista) lo que caracteriza a nuestra especie no es sólo (ni tanto) la capacidad de «inventar tecnología» como la de transmitirla, de forma acumulativa, de unas generaciones a otras. Esa necesidad requeriría, a su vez, la formación de grupos sociales complejos, caracterizados por: (a) un alto grado de estabilidad, (b) la vigencia prolongada de un periodo inmaduro, crítico para el aprendizaje, en las crías, (c) la convivencia y cooperación de miembros de distintas generaciones, (d) la existencia de periodos prolongados de actividad aparentemente «improductiva». En los grupos así constituidos, formados por organismos con notables capacidades de cálculo y anticipación de las conductas ajenas, y de acceso a los estados mentales propios, la «astucia psicológica» requerida habría sido cada vez mayor, al ser evolutivamente ventajosa la predicción —con fines cooperativos o competitivos— de la conducta de los congéneres, y la «penetración» en estados mentales de sistemas tan complejos como el propio sistema que predice o penetra en la interioridad de sus congéneres. En estas condiciones, los sistemas más capaces de (a) acceder introspectivamente a su propia complejidad y (b), tomándola como modelo, preveer más «jugadas» posibles en los juegos de interacción social, debieron contar con evidentes ventajas evolutivas.

Recientemente, Leda Cosmides (1989) ha proporcionado algunas pruebas favorables a una concepción semejante a la que defiende Humphrey, y obtenidas en estudios sobre razonamiento adulto con el conocido problema de las «cuatro tarjetas», inventado por Peter Wason (1966, 1968; Wason y Johnson-Laird, 1972). Como es sabido, el problema implica contrastar una hipótesis, establecida en una forma condicional: «si p entonces q», decidiendo, qué tarjetas habría que volver de cuatro que contienen, en su cara visible los casos «p», «q», «-p» y «-q», y aceptando

el supuesto de que toda tarjeta contiene en una cara un caso «p» (que puede ser «p» o «-p»).y en la otra un caso «q» («q» o «-q»). Por ejemplo, las tarjetas pueden contener en una cara una letra y en otra un número, y el enunciado condicional ser «Si una tarjeta tiene una vocal en una cara, entonces tiene un número par en la otra», mostrando las tarjetas presentadas, en su cara visible, los casos «E», «K», «4» y «7». La tarea del sujeto consistiría en decir qué tarjetas habría que volver para contrastar la verdad o falsedad del enunciado condicional.

Hay numerosas investigaciones que demuestran que, en esta forma abstracta, el problema de las cuatro tarjetas resulta extraordinariamente difícil para la mayor parte de los adultos, aunque la dificultad depende mucho del contenido, facilitándose cuando éste es más concreto y cercano a la experiencia del sujeto que razona (Bracewell y Hidi, 1974; Brown, Keats, Keats y Seggie, 1980; Cox y Griggs, 1982; Griggs y Cox, 1982, 1983; Johnson-Laird, Legrenzi y Legrenzi, 1972; Manktelow y Evans, 1979; Van Duyne, 1974; Wason y Shapiro, 1971; Yachanin y Tweney, 1982) y sobre todo cuando se suscitan «esquemas pragmáticos de razonamiento», dependientes de los contenidos (Cheng y Holyoak, 1985). Lo que ha demostrado Cosmides (1989) es que el problema de las cuatro tarjetas se facilita extraordinariamente cuando se establece en la forma de un problema de «intercambio social», que implica descubrir posibles «engaños», con respecto a una regla social.

Uno de los ejemplos de Cosmides (1989) puede servir para ilustrar el procedimiento que emplea: Al sujeto se le hace imaginar que él es un kaluama, un miembro de una cultura polinesia con reglas muy estrictas. Entre los kaluamas, cuando un hombre se casa, se hace un tatuaje en la cara. Los casados llevan tatuajes y los solteros no. Otra peculiaridad de la cultura kaluama es que hay ciertas reglas para consumir determinados alimentos: en concreto, la raíz de casava y las nueces de nolo. La raíz de casava es deliciosa y nutritiva, pero, sobre todo, tiene poderosas propiedades afrodisiacas, y otras propiedades tales que hacen irresistible a quien la consume. Las nueces de nolo saben bastante mal, no son muy nutritivas y no poseen esas propiedades. Dadas las estrictas costumbres de los kaluamas, los ancianos han establecido una ley, que establece lo siguiente: «Si un hombre come raíz de casava, entonces debe tener un tatuaje en la cara». El sujeto del experimento hace el papel de un guardián de un campamento provisional en que están cuatro kaluamas jóvenes: su misión consiste en descubrir a aquellos jóvenes que pudieran haberse saltado ta ley, ante la urgencia de sus impulsos sexuales. Cada

tarjeta representa a un joven, en una cara se señala si su cara está o no tatuada, y en otra lo que come; en una se lee: «come raíz de casava»; en otra: «no tatuado»;, en la tercera: «come nueces de nolo», y en la cuarta: «tatuado».

A diferencia de lo que sucede con contenidos más abstractos y no-sociales, la mayor parte (un 75 %) de los sujetos saben qué tarjetas hay que volver, en la situación anterior, para descubrir a los posibles «tramposos». Cuando el problema de las cuatro tarjetas, conservando su estructura formal clásica, se convierte en un problema de «intercambio social», que implica descubrir quiénes hacen «trampa», se produce un efecto claro de facilitación. Sirviéndose de una ingeniosa serie de experimentos en que empleaba diferentes variaciones del problema de las cuatro tarjetas, Cosmides (1989) ha defendido la idea de que la mente humana está especialmente «diseñada», por las fuerzas de la evolución *para resolver problemas interpersonales de intercambio social.* Propone que, cuando los mecanismos que definen la arquitectura cognitiva del razonamiento, se consideran desde una perspectiva evolucionista, la hipótesis de que tales mecanismos son generales e igualmente aplicables a todos los dominios resulta muy poco plausible.

En las condiciones pleistocénicas de la evolución humana, debió resultar más eficaz la evolución de mecanismos específicos (de *módulos* especializados de razonamiento, si se quiere decir así), que permitirían realizar, de forma rápida y eficiente, inferencias en situaciones de interacción e intercambio social.

Basándose en reconstrucciones de los factores evolutivos que presionaron la selección genética humana durante más de dos millones de años, y en los resultados experimentales de sus investigaciones, Cosmides (1989) propone la existencia de «algoritmos darvinistas de regulación del intercambio social» que: (a) producirían y procesarían representaciones de los costes y beneficios en las interacciones, y (b) permitirían realizar inferencias complejas sobre las posibles «trampas» en las situaciones de intercambio (i.e. detectar a los individuos que reciben los beneficios sin pagar los costes, como los solteros que comen raíz de casava, aprovechándose de sus propiedades afrodisiacas sin pagar los costes matrimoniales estipulados).

Ahora bien, parece evidente que la detección de la trampa, el engaño y la mentira (como la realización de trampas, engaños y mentiras) sólo es posible en *sistemas capaces de atribuir deseos y creencias,* es decir, en sistemas que, además de poseer mente, atribuyen mente. En el ejemplo

de la tarea a la que nos hemos referido, del experimento de Cosmides (1989), la actividad del sujeto tiene sentido en tanto que *sabe* que la raíz de Casava es un afrodisiaco tentador, y que los hombres —casados y solteros— *desean* comer raíz de casava, por lo que los solteros pueden tener la *intención,* (más en concreto, la *tentación*) de violar la ley que establece que «Si un hombre come raíz de casava, entonces debe tener un tatuaje en la cara», aunque esos hombres *conozcan* él contenido de esa ley.

Más aún: del mismo modo que la independencia estimular de la conducta y su determinación «de arriba a abajo» son los *criterios* básicos para determinar que un sistema es un sistema cognitivo, las competencias de realizar engaños y mentiras, y de detectar a aquellos que son engañados o engañan, de los que mienten o son equivocados por la mentira, es el criterio básico para admitir que un sistema tiene la capacidad meta-representacional de atribuir estados mentales.

Un organismo que, de forma deliberada, es capaz de transmitir a otro una creencia equivocada —que el primer organismo sabe que no corresponde a la realidad—, con el fin de obtener un beneficio, da muestras de ser capaz de: (a) atribuir a otro organismo estados mentales y (b) diferenciar los estados mentales ajenos de los propios. Posee, en sentido metafórico, una cierta «teoría de la mente».

Recordemos que Premack y Woodruff (1978) acuñaron este importante concepto de «teoría de la mente», para explicar la competencia de la chimpancé Sarah de elegir la «solución» correcta a ciertos problemas filmados que eran representados por humanos (por ejemplo, un humano enjaulado trataba de alcanzar un plátano en una película, y Sarah sabía elegir luego el instrumento que le permitiría alcanzarlo de entre otros varios, mostrados en tarjetas). Premack y Woodruff señalaron que, para dar solución a esos «problemas de otros», la chimpancé no sólo tenía que tener la capacidad de seleccionar la respuesta correcta, sino la competencia previa de discriminar que los humanos que aparecían en las películas «tenían un problema» y «deseaban resolverlo». Para ellos, esa competencia implicaría la posesión de una «teoría de la mente», i.e. de una capacidad de emplear conceptos no directamente empíricos (tales como son los conceptos mentalistas y muchos de los que se emplean en la ciencia) para predecir ciertos fenómenos (del mismo modo que se predicen mediante los conceptos científicos), en este caso, fenómenos de conducta.

Algunos de los investigadores que comentaron el importante artículo

de Premack y Woodruff (1978), publicado en *The Behavioral and Brain Sciences*, señalaron que el criterio más estricto y correcto para demostrar que un organismo posee realmente la capacidad meta-representacional de «tener representaciones sobre las representaciones de otros», es decir, de «tener una teoría de la mente», sería la demostración de su competencia para engañar deliberadamente.

La posesión de esta competencia por chimpancés ha sido demostrada experimentalmente por Woodruff y Premack (1979) y registrada en situaciones naturales por Jane Goodall (1986), entre otros. Un ejemplo de Goodall es el siguiente:

> En una ocasión, y cuando un grupo había terminado su comida, Figan descubrió un plátano que los demás habían pasado por alto, pero desgraciadamente Goliat se encontraba descansando debajo de la rama en que se hallaba. Aquél, después de dirigir a su congénere una rápida ojeada, se instaló al otro lado de la tienda, en un lugar donde no podía ver el plátano. Cuando, quince minutos después, Goliat se levantó para irse, Figan, sin dudarlo un segundo, se dirigió al árbol y tomó la fruta. Evidentemente había comprendido la situación; si hubiera pretendido coger el plátano bajo la vigilancia de Goliat, éste se lo habría arrebatado, y si se hubiera quedado donde estaba, no hubiera podido evitar lanzar alguna ojeada a la codiciada fruta, lo cual, dada la facilidad de los chimpancés para interpretar las miradas de sus compañeros, habría puesto a Goliat sobre la pista. Por lo tanto, Figan no sólo había conseguido sus deseos, sino que se había apartado del lugar peligroso que podía delatarle. Hugo y yo quedamos, naturalmente, muy impresionados, pero Figan había de sorprendernos aún más en otras ocasiones (p. 83).

Es fácil ver la delicada y compleja maquinaria inferencial que se pone en juego en el ejemplo anterior: Figan demuestra una capacidad considerable como «psicólogo natural» mentalista. Sabe que, si no puede resistir la tentación y trata de hacerse con el plátano, Goliat —que por algo se llama como se llama— hará valer su fuerza. *Calcula* que si él cae en una tentación más sutil, y se limita a mirar el plátano, proporciona a Goliat una pista peligrosa. Lo que hace entonces, es alejarse y esperar a que Goliat se marche, para conseguir su merecido premio (pues la habilidad «psicof lica» de Figan bien merece, cuando menos, un plátano). Pero esto no es quizá suficiente, para atribuir una teoría de la mente activa a Figan, pues, más que «engañar» a Goliat, lo que hacía era ocultarle pistas. Veamos entonces un ejemplo más claro:

Por lo general, cuando los chimpancés han estado descansando, si uno de ellos se pone en pie y emprende la marcha, los demás le siguen inmediatamente /.../. Un día en que Figan, por acompañar a un grupo numeroso, no había podido conseguir más de un par de plátanos, se levantó súbitamente y empezó a caminar. Los otros le imitaron. Diez minutos después regresaba al campamento él solo y recogía, libre de competencia, su ración de plátanos. Pensamos que se trataba de una coincidencia /../ pero, cuando repitió la misma maniobra una y otra vez, no tuvimos más remedio que aceptar que lo hacía deliberadamente (p. 84).

Como vemos en los ejemplos anteriores, las habilidades psicológicas, que presupone la atribución de estados mentales a otros, no son exclusivas de los humanos. Se dan en antropoides superiores, en chimpancés y gorilas y (quizá en menor grado) en orangutanes. Estos animales también son, en cierto sentido, «psicólogos mentalistas». Sin embargo, sus capacidades psicológicas son limitadas cuando se comparan con las que tienen niños de pocos años. En los últimos años se han realizado numerosas investigaciones evolutivas sobre la «teoría de la mente», que han demostrado que los niños de cuatro años y medio son capaces de distinguir con claridad entre su propio conocimiento de un estado de hechos objetivo y la creencia (equivocada) que atribuyen a un personaje, o a otro niño, acerca de esa situación (Astington, Harris y Olson, 1988). En uno de los modelos experimentales empleados, se muestran al niño dos personajes, que están en una habitación. Cada uno tiene un recipiente (por ejemplo, A un bolso y B una caja), y uno de ellos posee un objeto valorado en su recipiente (A tiene en el bolso una canica). Este personaje abandona la habitación, y el otro «se hace» con el objeto valioso y lo pone en su propio recipiente (A se marcha y B pone la canica en su caja). Después vuelve el primer personaje, y se pregunta al niño *dónde creerá* que está su canica, y *dónde la buscará.* Los niños normales de cinco años suelen ser capaces de decir correctamente que A buscará el objeto en su bolso, y creerá que está en él, aun cuando ellos mismos saben que está en el otro recipiente.

Lo que resulta más sorprendente de la capacidad infantil de diferenciar las propias creencias de las que tienen otras personas o personajes es el hecho de que se desarrolla mucho antes de lo que sería predictible si partiésemos del supuesto de que los niños emplean *competencias cognitivas generales* para resolver problemas de «teoría de la mente». El ejemplo del problema de A y B puede servir para aclarar este punto: saber que, si sucede que A tiene una creencia *«x»,* que corresponde a un estado de cosas X y tal estado de cosas se transforma en otro, Y, de

forma que A no percibe la transformación de X a Y,(aunque el propio niño sí la perciba), entonces A mantendrá su creencia «x», parece —cuando menos— tan complicado como saber que si en dos vasos se pone la misma cantidad de líquido y luego se pasa el líquido de un vaso a otro recipiente, seguirá habiendo el mismo líquido, aunque «parezca» más alto.

Este segundo problema, que es uno de los paradigmas de los problemas «operatorios» de conservación de Piaget, se resuelve mucho más tarde que tareas de teoría de la mente que también parecen exigir ciertos niveles de «reversibilidad» y «conservación» (Rivière y Castellanos, en preparación).

En suma: hay razones para pensar que, del mismo modo que la compleja tarea de las cuatro tarjetas se facilita enormemente cuando se convierte en un problema de «intercambio social» y reconocimiento del engaño, también se facilitan tareas de carácter operatorio cuando se establecen en forma de problemas de «teoría de la mente».

7.3. **Psicología natural, intensionalidad** y **metarrepresentación**

Las concepciones acerca del pensamiento y el razonamiento humanos que parecen derivarse de estas investigaciones recientes se apartan, de forma significativa, de los conceptos predominantes durante muchos años acerca de estas funciones. Hay fundamentos empíricos y argumentos teóricos suficientes para defender la idea de que *la mente humana está especialmente preparada para pensar acerca de la mente misma.* Dispone, probablemente, de mecanismos especializados de cómputo (quizá pudieran llamarse «módulos», en un sentido más amplio que el que emplea Fodor, 1983) para desarrollar sus inferencias mentalistas. En esas inferencias, emplea «representaciones sobre representaciones» de otras mentes o de sí misma. De tal modo que podemos decir que la mente humana posee una competencia *meta-representacional,* que proporciona el fundamento en el que se basa el componente pragmático del lenguaje y que está íntimamente relacionada con su capacidad *simbólica* (Leslie, 1988).

Alan Leslie (1987, 1988) ha desarrollado un modelo cognitivo muy sugerente de las capacidades subyacentes al desarrollo de la «teoría de la mente» en el niño. Su modelo parte de la observación de que existe un isomorfismo profundo» entre los juegos de ficción, que los niños empiezan a hacer desde el segundo año de vida, y las atribuciones men-

talistas que se efectúan en las tareas de teoría de la mente. Los juegos de ficción implican también una especie de «desdoblamiento de las representaciones»: la niña pequeña que emplea una cucharita para simular que habla por teléfono, lava la cara de su muñeca «como si» estuviera sucia, o «hace que come» con las manos vacías, está, como si dijéramos, «poniendo entre paréntesis» las representaciones realistas y «literales» de las cosas y configurando un mundo de representaciones ficticias, en el que se distorsionan deliberadamente las relaciones de referencia, verdad y existencia que tienen, con respecto a la realidad, las representaciones «verídicas» de las cosas.

Los mecanismos de *sustitución de objetos,* *atribución ficticia de propiedades* y *simulación imaginaria de objetos* (Rivière, 1990), que juegan un papel esencial en el desarrollo de las metáforas infantiles del-juego simbólico, muestran un curioso isomorfismo con ciertas propiedades importantes, de naturaleza semántica, de los enunciados que expresan estados mentales. Estas propiedades son las siguientes:

(1) Los enunciados intensionales son *referencialmente opacos.* Es decir, cuando dos expresiones con el mismo referente aparecen en contextos extensionales idénticos, si es verdad la primera, lo es la segunda; pero esto no sucede en determinados contextos con verbos mentales. Por ejemplo, del hecho de que sea verdad que «El presidente del Real Madrid veranea en Mallorca» se sigue que es verdad que «Ramón Mendoza veranea en Mallorca», pero de la verdad de que «Pedro piensa que el presidente del Real Madrid veranea en Mallorca» no se sigue la verdad de que «Pedro piensa que Ramón Mendoza veranea en Mallorca» (puede suceder, por ejemplo, que Pedro no sepa que Ramón Mendoza es actualmente el presidente del Real Madrid).

(2) Técnicamente, se dice que los contextos mentalistas con verbos intencionales son *intensionales* (en realidad, la intensionalidad es una propiedad que poseen, sobre todo, los verbos de pensamiento y lenguaje, tales como creer, interpretar, comprender, decir, suponer, intuir, etc). La opacidad referencial puede comprenderse como una consecuencia de la intensionalidad. Esta se refiere, principalmente, a la propiedad de «no-compromiso con la verdad o falsedad», a la propiedad por la cual la verdad de los enunciados completos no implica la de las cláusulas sometidas a los verbos mentales. Así, aunque sea verdad que «Luis cree que María está locamente enamorada de él», ello no implica que sea verdad que María esté realmente enamorada de Luis. Puede ser verdad la primera proposición sin serlo la segunda. Por el contrario, cuando un

enunciado extensional completo es verdadero, lo son todas sus cláusulas.

(3) La tercera propiedad que poseen los enunciados intencionales con verbos de pensamiento y lenguaje es el hecho de que su verdad no compromete la existencia de los objetos incluidos en cláusulas sometidas a verbos mentales de ese tipo. Así, puede ser verdad que «Antonio cree que el rey de Francia es calvo» y, sin embargo, no existe un objeto tal como el rey de Francia. Sin embargo, si es verdadero el enunciado extensional «El rey de Francia es calvo», ello implica que existe un objeto x tal que x es el rey de Francia.

Para Leslie (1987, 1988) existe un paralelismo claro entre la opacidad referencial de los contextos intensionales y el mecanismo simbólico de sustitución de objetos, entre el no compromiso con la verdad o falsedad de aquellos y la atribución simbólica ficticia de propiedades, y finalmente entre el no-compromiso de existencia de los objetos incluidos en cláusulas intencionalmente verdaderas y la simulación imaginaria de objetos. Explica ese paralelismo sirviéndose de la idea de que «la enunciación de estados mentales y la ficción son actividades que dependen cognitivamente de la misma forma especializada de representación mental y que, por consiguiente, *heredan sus propiedades semánticas*» (1988, p. 27). La forma específica de representación a la que se refiere Leslie es la metarepresentación, que implica un «desdoblamiento» («decoupling») de las representaciones primarias (perceptivas) y permite tanto la constitución de un mundo simbólico como la atribución simbólica de estados mentales a otros.

7.4. La conciencia como sistema de predicción mentalista y la semántica de los verbos mentales

La formulación de Leslie (1987, 1988) tiene la ventaja de que permite entrever las profundas, y aún mal comprendidas, relaciones que existen entre los mecanismos de cómputo que constituyen la base de las elaboradas competencias de interacción de los miembros de nuestra especie y los sistemas humanos de símbolos.

Desde una perspectiva genética, los símbolos son, como quería Vygotski (1979), formas de interacción, y su adquisición, forma, uso y desarrollo sólo puede explicarse (en el caso del hombre) como producto de una especie formada, si se permite esta expresión, por «psicólogos mentalistas». Los sistemas simbólicos permiten, a los humanos, acceder

al mundo interno y representacional de los congéneres mediante procesos muy elaborados y eficientes. Además, terminan por constituirse en el tejido con el que se elabora la conciencia reflexiva humana, cuya estructura semiótica está al servicio de su función esencial de «comunicación con uno mismo». Pero, por otra parte, no sería explicable la adquisición de esos sistemas simbólicos si no fuera por la posesión previa de mecanismos más primitivos de acceso al mundo interno propio y ajeno.

Todo esto significa que la psicología natural mentalista que, como veíamos en el capítulo anterior, constituye un punto de partida inevitable (aun cuando ésta termine por negar su valor) para el desarrollo de la psicología científica, es sencillamente una de las propiedades esenciales del sistema cognitivo humano.

Según este punto de vista, la conciencia reflexiva, tan difícilmente asimilable por los modelos computacionales de la mente, puede concebirse como un instrumento muy eficaz de representación acerca del mundo propio: un instrumento que estaría al servicio de la comprensión mentalista y la predicción de la conducta de los congéneres. La simulación representada de los objetos y procesos internos de éstos, mediante el empleo del plano de la mente proporcionado por el conocimiento de la mente propia, constituiría la estrategia principal de la psicología mentalista. Pero, por otra parte, esa psicología natural se vería limitada —en sus posibilidades de discriminación conceptual y, por consiguiente, de inferencia— a tratar con propiedades de lo psicológico que se dan en el plano de la conciencia.

Esta última observación es importante para comprender las complicadas relaciones que siempre ha tenido la psicología científica con la psicología natural. Si ésta es el resultado de un sistema (muy probablemente especializado y modular) de cómputo que se sirve de la conciencia propia y del supuesto de identidad entre congéneres, para comprender y predecir la conducta de otros, ello querrá decir que *todos los rasgos y propiedades de lo mental que se representan en el sistema conceptual de la psicología natural provienen de rasgos y propiedades que se manifiestan en la conciencia.* Y el problema es que la conciencia —como ha demostrado, de forma muy clara, la historia de la psicología— no parece estar constituida para informar sobre la naturaleza interna de las funciones mentales, sino más bien para definir significativamente los contenidos sobre los que éstas se realizan y sobre algunos aspectos de la relación del sujeto con tales contenidos.

Si nos fijamos, por ejemplo, en las diferencias semánticas entre los

verbos mentales de contenido epistémico, vemos que esas diferencias parecen corresponder a un conjunto limitado de primitivos conceptuales que reflejan aspectos conscientes de nuestras relaciones con contenidos mentales: aspectos tales como el «grado de certeza» o de «seguridad» con que se conocen esos contenidos (que se refleja, por ejemplo, en la diferencia entre «saber» y «creer»), el hecho de que se sometan a cambios o se mantengan estáticos («pensar» frente a «creer»), de que se sitúen en el presente, el pasado («recordar») o el futuro («prever», «anticipar»), o la gran distinción que hace el sistema cognitivo entre contenidos a los que asigna una relación inmediata, presente y casi-extensional o física con el medio («percibir») y otros que implican una distancia mayor y una elaboración inferencial más compleja («pensar», «intuir», «recordar», etc). En-suma, las dimensiones que organizan el campo semántico de los verbos mentales se derivan de propiedades de la conciencia de la relación con objetos a que estos verbos hacen referencia. Esa conciencia de relación es muy limitada, y no implica, ni mucho menos, que las funciones que permiten la formación, elaboración y transformación de los contenidos mentales sean *transparentes.* Por el contrario, esas funciones son esencialmente opacas a la conciencia.

Otra forma de expresar lo que significan las afirmaciones anteriores es la siguiente: las *actitudes proposicionales,* de que hablan los filósofos y que constituyen la materia prima de la psicología natural mentalista, no se identifican con las funciones mentales de que hablamos los psicólogos cognitivos. En tanto que estas funciones se conciban como algoritmos de cómputo que, en función de la *organización* del conocimiento en el sistema cognitivo, pretendan explicar tanto la conducta como los contenidos de la conciencia, no podrán identificarse con las actitudes proposicionales que tienen la función esencial de predecir la conducta de otros sirviéndose de un supuesto de similitud entre los aspectos limitados del mundo propio, a los que la conciencia tiene acceso, y los mundos internos de los congéneres.

Las actitudes proposicionales no definen *mecanismos de producción* de los contenidos de la mente, sino formas de relación entre el sujeto y tales contenidos. La gama limitada de formas de relación que se refleja en las actitudes proposicionales es también el producto de una elaboración evolutiva: podemos suponer que, a lo largo de la evolución humana, resultaron adaptativas capacidades que permitían: a) diferenciar claramente los objetos percibidos de las elaboraciones del pensamiento y el recuerdo; b) discriminar aproximadamente el grado de probabilidad o

seguridad asignado a las creencias; c) asignar valores temporales a los contenidos representacionales de la conciencia; e) distinguir las valencias de los objetos, etc. Estas propiedades, proyectadas en la conciencia, son las que se reflejan en el vocabulario de las actitudes proposicionales.

Del mismo modo que, como decíamos en el capítulo primero, nuestro sistema cognitivo parece estar adaptado a un «mesocosmos» físico, constituido por objetos de mediano tamaño y al que se accede a través de la gama limitada de energías del medio que transducen nuestros receptores sensoriales (de manera que las ciencias físicas han tenido que superar las restricciones y sesgos que establecen esas adaptaciones que han resultado de nuestra evolución específica), así también nuestra «psicología natural» parece estar adaptada a un «mesocosmos» psicológico que depende de un conjunto muy limitado de discriminaciones a las que se accede por los sistemas de autobservación e introspección ingenua que han resultado de la evolución peculiar de una especie social y psicológica como la nuestra. Y de la misma forma que la física científica no puede permanecer anclada en el mesocosmos al que está preadaptado el sistema cognitivo humano, la psicología científica no puede estancarse en la concepción «media» de la mente que sirve de base a la psicología natural.

7.5. Psicología natural y psicología popular

Es importante diferenciar esa psicología natural de lo que algunos investigadores han llamado «psicología popular», es decir de las «explicaciones» sobre lo mental —por muy infantiles y arcaicas que éstas sean— que se realizan fuera de los ámbitos científicos y que varían en las distintas sociedades y culturas. La indistinción entre una y otra, entre psicología natural y psicología popular, ha dado lugar a equívocos filosóficos considerables (López Cerezo, 1989; De Vega, 1989, Rivière, 1989). Como ha señalado De Vega (1989) una y otra constituyen dominios fenoménicos diferentes: el primero se refiere «al conocimiento pragmático e instrumental que posee el hombre de la calle acerca del entorno psico-social donde se desarrolla su interacción con otras personas» (p. 275). El otro dominio, el de la psicología popular, se deriva de la propensión teórica y explicativa inevitable que tiene el hombre, que «no solamente dispone de un conocimiento pragmático sobre los demás y de una metacognición sobre sí mismo, sino que, además, pretende establecer principios generales para *explicar* el mundo psicosocial y *justificar* su propio comportamiento» (ibídem).

Aunque se utilice el término «teoría de la mente» para referirse a los dominios de la meta-cognición y el conocimiento pragmático de los demás que hemos reflejado en este apartado, el concepto de «teoría» es metafórico (y peligrosamente ambiguo) cuando se aplica a la psicología natural. Las unidades de conocimiento de que se compone la psicología natural no son «formas de conocimiento declarativo, sujeto a valores de verdad, públicamente contrastable, etc.» (De Vega, 1989, p. 276). La diferencia entre la psicología natural y la psicología científica es, así, semejante a la que existe entre el conocimiento pragmático tácito y el conocimiento explícito de la gramática de un lenguaje. Del mismo modo que *sabemos* pragmáticamente las personas acerca de la mente, *sabemos* hablar y comprender. En nuestros usos ordinarios del lenguaje, ponemos en juego estructuras de conocimiento muy complejas, tales como las que nos permiten distinguir los rasgos fonológicos relevantes en nuestro sistema lingüístico, comprender y producir potencialmente infinitas oraciones gramaticales y llevar a cabo sutiles distinciones semánticas. Así, *conocemos,* en cierto sentido pragmático, la sintaxis, la fonología y la semántica de nuestro lenguaje. Pero eso no quiere decir que podamos establecer de forma explícita sus reglas fonológicas, sintácticas y semánticas. En realidad, esas reglas son tan complejas que no han podido ser reconstruidas completamente en ningún lenguaje natural, a pesar de los esfuerzos continuados de los lingüistas.

Algo semejante sucede con el dominio, de lo mental, que es el campo al que los psicólogos dedicamos nuestros esfuerzos profesionales: la psicología natural es, como el lenguaje, un refinado producto evolutivo de la filogénesis y la ontogénesis (quizá también de la «sociogénesis») de la mente humana. Entre las propiedades adaptativas que ésta presenta, como producto contingente de la evolución, hay una importante: la que consiste en diferenciar y resaltar las representaciones acerca de sí misma y acerca de otras mentes. Es ésa una propiedad explicable en un animal social y lleva a la construcción de sistemas específicos de computo para comprender y predecir conductas humanas; sistemas en que la conciencia, y los mecanismos metacognitivos que implica, juegan un papel decisivo.

De este modo, la intencionalidad humana presenta una importante y poco analizada propiedad que no parecen tener las funciones intencionales de otros animales, y que parace fundamental para entender algunos de los aspectos importantes del pensamiento y el lenguaje. Me refiero a la propiedad de ser recursiva (Rivière, 1990): las funciones mentales son

funciones que consisten en relaciones de sujetos con objetos, pero que tienen la propiedad de que tales objetos pueden ser, a su vez, funciones mentales. Cuando en el texto de Ross MacDonald se decía: «Hillman no dijo nada. Tal vez supiera algo de su hijo que los otros ignoraban», encontramos que se atribuye a Archer un estado mental (pensar algo) que se refiere, a su vez, a otro (saber algo) atribuido a otro personaje (Hillman), siendo ese algo un objeto que se relaciona, a su vez, con un tercer verbo epistémico (ignorar algo). Como en el juego de las muñecas rusas, las actitudes proposicionales pueden encajarse una en otras (i.e. poseen un mecanismo de auto-incrustación) y ramificarse unas de otras (ramificación), creando estructuras de tipo (I_1 (I_2 (I_3))).

Una característica esencial de los miembros de nuestra especie puede ser esa *intencionalidad recursiva* (empleando aquí el término «recursivo» en un sentido no-técnico, que debe diferenciarse claramente de su significado matemático preciso, tal como se define, por ejemplo, en Arbib, (1964)). Es decir, esa capacidad de atribuir a otros estados mentales intencionales que, a su vez, pueden tener como objetos estados mentales intencionales. En *Linguistic Behavior,* Bennett (1976) defiende la idea de que la adquisición de un sistema simbólico tal como el que constituye el lenguaje humano requiere, cuando menos, una intencionalidad de tercer orden (algo similar a «saber que el otro sabe que yo sé que...»). A pesar de que los antropoides superiores alcanzan un segundo orden de intencionalidad («creer que el otro cree que...»), como han demostrado Premack y Woodruff (1978), y Woodruff y Premack (1979), no existen pruebas que permitan afirmar que alcancen esa intencionalidad recursiva de orden superior que se pone en juego, constantemente, en los usos lingüísticos de los humanos.

Su muda, silenciosa, atribución de intenciones se queda, probablemente, a medio camino entre la mera posesión de estados mentales intencionales (que puede predicarse, en alguna medida, de todos los animales que tienen estados epistémicos, ya que todo estado epistémico es, por definición, intencional) y la intencionalidad, ya claramente recursiva, que se encuentra en el hombre.

Del mismo modo que los psicólogos cognitivos tenemos la obligación de explicar qué mecanismos de cómputo permiten al hombre crear mundos visuales tridimensionales, comprender oraciones gramaticales y jugar al ajedrez o resolver silogismos lineales, también deberíamos enfrentarnos a la tarea de explicar los mecanismos de cómputo que permiten configurar un mundo social basado en los supuestos de intencionalidad

recursiva e identidad esencial entre la estructura de las mentes, y en la atribución de estados mentales a los congéneres. La propensión mentalista de nuestro sistema cognitivo es tan profunda que tendemos a predicar mente de objetos físicos, cuyo funcionamiento puede explicarse sin el supuesto de que manejan representaciones internas. El animismo, al que nos hemos referido en el capítulo sobre los objetos con mente y la metáfora del ordenador, puede entenderse, desde esta perspectiva, como una propensión derivada de la especialización mentalista del sistema cognitivo.

La psicología natural incluye, en su núcleo, un cierto modelo implícito del sujeto humano. Ese modelo se articula alrededor de los conceptos de persona e intención: las personas son agentes intencionales, con capacidades de auto-conocimiento, auto-supervisión y auto-intervención (Harré, 1979). Son, por antonomasia, los sujetos de atribución de las actitudes proposicionales. Se entienden como entidades unitarias, y causas endógenas de la conducta. En las interacciones habituales de nuestra cultura, a esas entidades a las que llamamos personas les es imputable, al menos hasta cierto punto, la «responsabilidad» de sus actos. Tales actos se codifican, desde la psicología natural, como resultantes de intenciones, conocimientos y deseos, que se relacionan de forma compleja con las condiciones ambientales para dar lugar a la conducta. Las personas, que constituyen unidades básicas de la psicología natural, son aquellos seres que se atribuyen a sí mismas enunciados intencionales en primera persona, en la forma: «Yo Fm (p)». «Yo (creo/siento, etc) algo».

La competencia de formular enunciados intencionales en primera persona de singular («yo (pienso, creo, deseo, siento, etc) que X») parece ser reflejo de un sistema cognitivo, la conciencia reflexiva, que seguramente tiene una función esencial en la capacidad de producir esos mismos enunciados en tercera persona, en la competencia de atribuir a otros estados mentales. La psicología científica se ha enfrentado siempre a la difícil y paradójica exigencia de situarse en un plano explicativo que, sin ignorar las intuiciones mentalistas de la conciencia, las supere mediante explicaciones objetivas. Ha soñado la compleja intención de establecer un conocimiento objetivo acerca de las raíces mismas de la subjetividad humana.

Capítulo 8

LOS ENUNCIADOS INTENCIONALES Y LA BUSQUEDA DE OBJETIVIDAD EN PSICOLOGIA

8.1. Mentalismo, animismo y evolución de la razón

Como resulta evidente por lo que hemos dicho en los dos capítulos anteriores, la psicología natural concibe implícitamente las funciones mentales como procesos o estados *internos*, que se realizan sobre *contenidos* o representaciones. La intencionalidad y la interioridad son, así, las dos características esenciales que damos a lo mental en nuestras interacciones cotidianas con las personas.

Atribuir mente equivale, en esa psicología cotidiana, a presuponer un sistema «que obra por conocimiento»: en función de lo que percibe, interpreta, infiere, reconoce, recuerda, cree. Además atribuir mente parece implicar «atribuirle» mente, es decir, tener un cierto acceso consciente a las relaciones que uno mismo tiene con los contenidos representados. El criterio esencial para saber si un organismo es capaz de atribuir mente a otro es su habilidad para diferenciar entre sus propios estados de conocimiento y los del otro organismo, que se'muestra especialmente en (a) las destrezas de engaño, y (b) las competencias para reconocer cúando alguien es engañado.

El hecho de atribuir mente a otro organismo o sistema de cómputo parece implicar, por tanto, las capacidades de segundo orden que se corresponden con las anteriores de primer orden: saber que uno mismo (o una tercera persona) sabe, infiere, cree, piensa, recuerda, etc. Todo parece indicar que la evolución de la conciencia reflexiva ha estado, en nuestra especie, estrechamente relacionada con el desarrollo de las habilidades psicológicas de tipo pragmático que permiten a los humanos ser animales perspicaces, que engañan y reconocen el engaño, se ade-

lantan a la conducta de otros y codifican esa conducta sirviéndose de un vocabulario conceptual mentalista.

Es sensata la hipótesis de que el desarrollo de la inteligencia humana, a lo largo de la filogénesis, pudo tener una relación clara con los mismos factores que hicieron de la especie humana una especie mentalista. La «espiral mentalista» debió estar estrechamente relacionada, en la evolución de nuestra especie a lo largo del Pleistoceno, con el tipo de procesos que produjeron esa competencia intelectual a la que Descartes denominaba «razón universal».

Esta afirmación merece una justificación, aunque sea breve: a diferencia de lo que sucede cuando el hombre se enfrenta con fenómenos naturales o con otros animales, la relación cooperativa o competitiva con otros miembros de la misma especie implica la necesidad de predecir y comprender, de forma muy rápida, el «funcionamiento» de sistemas cognitivos y mentales del mismo nivel de complejidad que el propio sistema empleado para conocer.

El versátil dinamismo de las interacciones y la complejidad de los sistemas en interacción (cuando estos sistemas son dos homínidos inteligentes de la misma especie) debieron ser factores que favorecieron la selección de variaciones adaptativas que implicaran un «plus», un «extra», de capacidad de procesamiento con relación a otros congéneres. Las capacidades «añadidas» de codificación e inferencia que permiten al hombre desbordar, con mucho, las limitaciones de la percepción ingenua del mundo físico y abstraer en él relaciones cada vez más profundas, pudieron tener su origen en la selección continuada de los miembros de la especie favorecidos, en sus posibilidades de supervivencia, por sus capacidades de cómputo mentalista.

La «espiral evolutiva» que dio lugar a una especie simbólica (Cassirer, 1945), y capaz de una «razón universal», pudo generarse en ese proceso de selección continuada de aquellos miembros de la especie capaces de «dar un paso más allá» en las interacciones y las atribuciones de funciones mentales, y en la construcción de «teorías» cada vez más complejas de la mente. Estas teorías implicaban, en sí mismas, el uso de conceptos inferenciales que, al tiempo que se alejaban, cada vez más, de las conductas inmediatas, permitían una predicción más precisa y a más largo plazo de los comportamientos de los congéneres, y una planificación más elaborada y previsora de las conductas propias.

De este modo, la evolución de una «razón sobre las razones de otros», de un instrumento predictivo acerca de las conductas de los congéneres,

habría sido un factor decisivo en la evolución de una razón capaz de despegarse cada vez más de las volubles apariencias de la percepción. Los individuos más capaces en sus predicciones mentalistas habrían sido también los más hábiles para construir «prototeorías» sobre el mundo físico, progresivamente más independientes de los fenómenos inmediatos. Si bien estas prototeorías fueron, en un primer momento, fuertemente animistas, expresaban ya el desarrollo de un sistema de conocimiento de naturaleza racional, que posibilitaría inferencias cada vez más audaces y compuestas de más pasos mentales. Probablemente, si en la evolución humana fueron seleccionados los individuos capaces de dar más pasos mentales, ello se relacionó más con las destrezas mentalistas que con las exigencias de conocimiento físico del mundo.

Ello explicaría la universalidad y profundidad de las propensiones animistas de explicación de lo físico en los niños y las culturas primitivas, al entender tales propensiones como distorsiones conceptuales resultantes de las influencias de factores sociales en la evolución de nuestra especie. Si la psicología natural de ella es mentalista, también lo es su «física natural»: durante miles de años, el hombre ha atribuido conocimientos, deseos e intenciones a los fenómenos naturales y, como comentábamos en el capítulo segundo, el desarrollo de la física científica sólo fue posible, al vencerse, en una larga y dura historia, las inercias animistas del pensamiento natural, en un profundo proceso de «desmentalización» del mundo físico.

8.2. Funciones mentales y objetividad de las explicaciones científicas

Ahora bien: «Desmentalizar» equivale, naturalmente, a lo contrario que «mentalizar», es decir, a dar un carácter *externo* y *estrictamente extensional* a los conceptos y enunciados sobre el mundo físico. ¿Qué significa exactamente dar un carácter externo y extensional a los conceptos?; ¿no es, por definición, un concepto algo interno e intencional?

Lo es, desde luego, como hecho psicológico. Pero una condición básica para el desarrollo de una ciencia *objetiva* es, precisamente, la exigencia de que los conceptos, proposiciones y enunciados dejen de ser hechos psicológicos, convirtiéndose en entidades que no son subjetivas, sino intersubjetivas (para algunos filósofos de la ciencia) o, más aún, suprasubjetivas (para otros). Con el vocabulario que hemos empleado hasta aquí, esto mismo puede expresarse de una forma muy simple: los

enunciados científicos objetivos no tienen la forma «X Fm (p)» («X piensa, cree, infiere, opina, etc, que p»), sino más escuetamente la forma «p» (así no es un enunciado científico «Juan piensa que un cuerpo en movimiento rectilíneo y uniforme, en el vacío, mantiene el movimiento», sino solo «un cuerpo...etc.»).

Es cierto que el origen de los enunciados «p» es siempre algún tipo de enunciado «X Fm (p)»: la génesis de los conceptos de las ciencias reside en el conjunto de procesos Fm, un conjunto de funciones mentales gracias a las cuales los científicos *perciben* ciertos fenómenos, *abstraen* relaciones entre ellos, *infieren* hipótesis y consecuencias de tales hipótesis, *piensan* teorías generales, etc. Sin embargo, estos procesos, que pueden ser relevantes para la explicación de los contextos de descubrimiento de las teorías científicas, tienen que ser *eliminados* de los enunciados para que éstas puedan situarse en contextos de justificación adecuados.

¿Cómo se efectúa ese proceso de eliminación de las funciones Fm que se incluyen originariamente en la actividad científica (de carácter «Fm (p)»), de tal manera que los enunciados científicos puedan tomar su forma definitiva, «p»? " -

Sencillamente, mediante los conocidos procesos de *objetivación* que se emplean, de forma habitual, en la práctica científica:

(1) En primer lugar, asegurando la replicabilidad de las observaciones en que se basa esa práctica, y el control estricto de las condiciones en que tales observaciones se realizan, de tal manera que pueda decirse no ya que «alguien percibe, observa, etc, (p)», sino que «sucede que (p)».

(2) En segundo lugar, separando, en lo posible, las observaciones de las inferencias.

(3) También asegurando el rigor y la universalidad y replicabilidad de las inferencias, mediante el uso de procedimientos lógicos y matemáticos, y el empleo de criterios de coherencia y precisión, de forma que pueda decirse no sólo que «alguien piensa que (p) se sigue de (q)», sino enunciarse que «p se sigue de q», etc.

Los métodos experimentales, las cautelas de observación, el uso —cuando es posible— de «procedimientos automáticos» de inferencia (como son, en general, los procedimientos matemáticos que implican el uso de algoritmos precisos y bien definidos en sus pasos), aseguran la justificación de enunciados en los que no figura ninguna referencia a procesos mentales de un determinado sujeto.

Así, una forma de explicar la razón por la que los científicos usan procedimientos complejos de observación e inferencia, y no se contentan con lo que observan a primera vista o lo que se les ocurre de buenas a primeras, es ésta: intentan, por todos los medios a su alcance, convertir los enunciados «X Fm (p)» en enunciados «p». Replican observaciones, las hacen públicas, emplean procedimientos de inferencia lo más precisos posible, controlan las condiciones de sus experimentos...y hacen todo ello con la firme voluntad de pasar de enunciados que contienen verbos mentales a otros que ya no los contienen.

De este modo, el desarrollo de la Ciencia Natural ha exigido históricamente dos conjuntos de procesos de *objetivación:*

(1) En primer lugar, procesos que podríamos llamar de *limpieza conceptual*', mediante los cuales se han eliminado las explicaciones animistas de los fenómenos, sustituyéndolas por otras de carácter mecanicista o, en cualquier caso, no animista (así, los cuerpos no se aceleran al acercarse al centro de la tierra porque se «alegren» de alcanzar su lugar natural, como se pretendió en algún momento, sino debido a las leyes de gravedad).

(2) En segundo lugar, procesos de *limpieza metodológica*, que permiten eliminar los verbos mentales de los enunciados observacionales y teóricos de las ciencias, por los procedimientos ya indicados: repetición y control de observaciones, automatización de procesos de inferencia, etc.

Los procesos de (a) «desanimación» de los fenómenos naturales y (b) objetivación de los enunciados científicos (es decir, de eliminación'de la atribución de funciones Fm a los fenómenos naturales, por una parte, y de cláusulas Fm en los enunciados observacionales o teóricos de la ciencia, por otra) han sido históricamente procesos estrechamente entrelazados, pero que deben distinguirse cuidadosamente.

Los primeros permitieron desarrollar una concepción mecanicista de la naturaleza (a la que ya nos hemos referido en el capítulo segundo), en la que los fenómenos no se comprendían como manifestaciones de sustancias o mentes sino, a través de sus relaciones, como especificaciones de leyes de interacción precisas entre las propiedades puramente extensionales de la materia y la energía. Los segundos procesos permitieron la justificación de las leyes formuladas mediante criterios relacionados con el carácter supra-subjetivo tanto de las observaciones como de las inferencias teóricas.

El efecto conjunto de los dos tipos de procesos fue el abandono

progresivo de las visiones antropocéntricas de la naturaleza, que siempre habían estado ligadas, de forma profunda, al empleo de los sistemas conceptuales especializados en la interacción con personas (es decir, de la psicología natural mentalista, utilizada para tratar con los prototipos de los objetos Fm) para dar cuenta de fenómenos y objetos que no requerían de la hipótesis de la mente para ser explicados.

Estas rápidas pinceladas sobre el desarrollo de las ciencias de la naturaleza no son más que una nueva versión de un viejo tema, desarrollado repetidamente por los historiadores y los filósofos de la ciencia. La ventaja que tiene, para nosotros, esta versión concreta es que nos permite describir con más claridad uno de los «dramas» históricos fundamentales de la psicología. Veamos en qué consiste ese drama:

Si las condiciones de posibilidad de las ciencias objetivas de la naturaleza residieron en la eliminación de los predicados Fm de los fenómenos naturales y, más aún, de los enunciados observacionales y teóricos de la ciencia que versa sobre ellos, eliminando así el «efecto del sujeto» que se enfrenta a tales fenómenos, ¿cómo es posible una ciencia que trata de ser objetiva, y de serlo, precisamente, sobre las funciones mentales?; ¿cómo es posible una ciencia objetiva de la mente?

8.3. Los enunciados con verbos mentales como obstáculos epistemológicos

Para dar respuesta a estas preguntas, es preciso que nos enfrentemos antes a una cuestión previa: ¿Por qué los enunciados que incluyen predicados Fm son especialmente difíciles de manejar en una ciencia objetiva de la naturaleza (como trata de ser la propia psicología)?

Las respuestas a esta última, y vital, pregunta pueden deducirse de algunas observaciones hechas hasta aquí. Sin embargo, merece la pena que nos enfrentemos de forma ordenada y explícita a nuestra pregunta: ¿qué dificultades plantean los enunciados Fm como proposiciones científicas?

(1) La primera dificultad está relacionada íntimamente con la inevitable vinculación a la conciencia de los enunciados mentalistas de la psicología natural. Es la siguiente: Prototípicamente, las funciones mentales se predican —ya lo hemos dicho— de las *personas.* Sólo las personas predican, *en primera persona de singular,* funciones mentales de sí mismas. Y sucede que los enunciados que incluyen esos predicados poseen una curiosa y evidente propiedad, a saber: cuando se enuncian en primera persona de singular no remiten ai mismo fundamento observacional (es decir, no tienen la misma base semántica; no tienen, en cierto modo, el mismo significado) que cuando se enuncian en segunda o tercera persona. Cuando un sujeto dice «Yo Fm (p)», «Yo (pienso, creo, intuyo, etc) que p», remite a un mundo que no es, en sí mismo, exteriorizable más que por sus *síntomas.* Hace un enunciado que, por así decirlo, *nadie, que no sea él mismo, puede corregir.* Un enunciado recluido, incorregible.

Nadie cuenta con el mismo fundamento observacional (con independencia de lo débil que éste pueda ser) que posee el sujeto que dice: «Yo Fm (x)» (por ejemplo, «yo creo o pienso que X»), como para permitirle corregir ese enunciado, diciendo: «No, Tú lo que haces no es Fm, sino Fm'», o «No, Tú no Fm (p), sino que Fm (q)». Así, los enunciados Fm tienen la peculiar característica de que son epistémicamente asimétricos: poseen propiedades en primera persona de singular que no tienen en segunda o tercera persona.

(2) Sin embargo, los enunciados científicos tienen que ser —si se admite este vocabulario— *estrictamente simétricos* (no *asimétricos)* para todas las personas, si quieren ser objetivos. La posibilidad de hacer inter-subjetivos y replicables los enunciados observacionales depende decisivamente de que su fundamento semántico sea *el mismo* para todos. De que sea universal, y no particular; intersubjetivo y no subjetivo; abierto y no clausurado. La intuición básica de la psicología natural, según la cual *los enunciados Fm se justifican, de forma esencial y prototípica, cuando se hacen en primera persona de singular,* choca frontalmente con las exigencias de objetivación de los enunciados científicos.

Estos tienen que ser corregibles, abiertos a la posibilidad de inspección de su base observacional, replicables. Por el contrario, cuando un sujeto dice: «Yo Fm (p)», lo que dice no puede corregirse, ni inspeccionarse con los mismos recursos observacionales en que se basa el enunciado en primera persona, pues nadie puede hacer introspección en la mente de otro. Así, los enunciados mentalistas en primera persona de singular han presentado siempre una paradójica ambivalencia en psicología: por una'parte, está la pregnante y poderosa intuición, que tenemos como psicólogos naturales, de que *sin ellos la psicología no sería posible, ni podría concebirse-,* la intuición de que, en cierto sentido, *la primera*

persona de singular es el «núcleo básico», la «residencia natural» de los enunciados psicológicos.

Por otra parte, y en tanto que tales enunciados no sólo articulan lo que entendemos por subjetividad sino que se encierran en ella, no ofrecen garantías como enunciados observacionales en los que pueda basarse la actividad científica. Así, los enunciados intencionales, o mentalistas, en primera persona de singular han sido, históricamente, fundamentos —implícitos o explícitos, pero inevitables— de la psicología científica y, al mismo tiempo, *obstáculos epistemológicos* (Bachelard, 1948) de primer orden para la constitución de una verdadera ciencia de la mente.

(3) Su carácter de obstáculos epistemológicos se acentúa aún más cuando se toma en consideración qué es lo que dicen, en realidad, esos enunciados, y qué lo que callan o ignoran.

Hemos visto, a lo largo de este capítulo, que los conceptos mentales con los que trata la psicología natural están completamente impregnados de conciencia, y limitados a la gama restringida de discriminaciones que se ofrecen a la fenomenología de la conciencia. Tales discriminaciones no se refieren al funcionamiento menudo de la maquinaria que produce eso a lo que llamamos mente, sino a aspectos relacionados con los objetos y contenidos sobre los que la mente versa.

La conciencia es, sobre todo, conciencia de algo y de mis relaciones con ese algo (me permite decir si el algo es ahora o ha sido, si me parece seguro, probable o improbable, si me atrae o repugna, si puedo o no afirmar que tiene, aquí y ahora, una consistencia extramental, etc.). Ofrece la engañosa imagen de que la mente es una maquinaria traslúcida, que «se ve funcionar a sí misma», pero sólo «ve» los objetos finales y las relaciones finales que resultan de las interrelaciones entre representaciones que se escapan, por completo, a sus posibilidades de observación introspectiva.

8.4. Las alternativas de tratamiento científico de los enunciados con verbos mentales

(1) La conciencia es así, principalmente, noticia de relación con contenidos y objetos, y no tanto registro de los procesos que permiten tal relación. Los primeros psicólogos experimentales, como Wundt y James, tenían ya muy presente el peligro de confundir la conciencia de contenidos con los contenidos de conciencia, pero aún conservaban la ilusión

de poder acceder introspectivamente al funcionamiento del sistema mental. Sin embargo, sus mejores descripciones (como la muy conocida de James, en los *Principles,* 1890) no pudieron ir más allá de lo que permitía una elaboración reflexiva de los principios de la psicología natural.

En esas descripciones el peso de la fenomenología de la conciencia era aún demasiado gravoso, y terminaron por producir la impresión de que la aparente trasparencia de la mente a su propia conciencia ocultaba, en realidad, una engañosa opacidad. El desánimo de James, al terminar los **Principles of Psychology** (al que nos referíamos en el capítulo primero) no era más que un reflejo de esa impresión, con la que se recapitulaba toda una época de la historia de la psicología.

(2) El abandono de los enunciados intencionales en primera persona de singular (i.e. de los enunciados introspectivos) como fundamentos observacionales de la psicología científica formó parte de la estrategia principal de las psicologías llamadas «objetivas» (especialmente la reflexología y el conductismo), en una primera fase. Ese abandono supuso un importante trauma histórico en psicología que no revisaremos ahora.

Ese movimiento histórico objetivista implicó, sin embargo, un paso más allá que resulta más difícil de justificar. Ese paso consistía en lo siguiente: supongamos que prescindimos, efectivamente, de proposiciones tales como «Yo (pienso que/creo que/siento que) x» en psicología. Imaginemos que sólo podemos decir: «El piensa/cree/siente, etc. X». Aunque ahora los enunciados no son en primera persona, siguen siendo enunciados *intencionales* y que atribuyen funciones *internas* a los sujetos. Pues bien: el segundo paso del objetivismo no-mentalista, que trataron de dar los conductistas, los reflexólogos, etc, consistió en intentar, por todos los medios, reducir los enunciados intencionales a enunciados extensionales, no solo en el lenguaje observacional sino también *en el lenguaje teórico* de la psicología.

En esta perspectiva, decir «El piensa que x» *significa,* por ejemplo, lo mismo que decir «El tiene tales o cuales reflejos inhibidos en el cerebro» (Sechenov), «El presenta tales o cuales respuestas de habla subvocal» (Watson), «El presenta conductas que se relacionan funcionalmente de tal o cual forma con el medio» (Skinner), «El encadena estímulos y respuestas mediacionales de tal o cual manera» (Osgood, Staats, Mowrer, etc). Los reflejos inhibidos, las respuestas en sus dimensiones cuantitativas, el habla, las relaciones funcionales entre estímulos y conductas *no* son objetos intencionales, ni poseen la clase de interioridad que tienen las funciones mentales de la psicología natural. Son, en todo

caso, fenómenos *potencialmente exteriorizables* y que tratan de ser *extensiofiales,* de tener un significado que no «apunta» hacia afuera de los propios fenómenos.

Estas interpretaciones de lo psicológico trataban de «replicar» en psicología el mismo proceso de «desanimación» y «desmentalización» que dio lugar a una ciencia mecanicista y estrictamente extensional de la naturaleza. Partían, generalmente, de una actitud suspicaz con relación a los conceptos teóricos, justificados de forma inferencial, de las teorías científicas. Al acentuar el componente observacional de la actividad científica, tendían a presuponer que las entidades teóricas que se incluyen en los modelos explicativos de los psicólogos deben consistir también en fenómenos potencialmente observables. Además, tales entidades no poseerían la nota de *intencionalidad* que poseen los conceptos mentalistas de la psicología natural y se relacionarían mediante asociaciones mecanicistas «de bajo nivel» (como las que se describen en los modelos de condicionamiento).

(5) Los modelos de lo mental así configurados eran, en cierto sentido, «contramodelos» de lo que es la mente para la psicología natural. No puede decirse que el sistema psicológico descrito por Watson o Skinner se caracterice por «obrar en función del conocimiento». Las notas de racionalidad, organización interna, intencionalidad, teleología y eficacia causal de la conciencia, por las que se define implícitamente el sujeto de atribución de la psicología natural no tienen fácil cabida en la mayor parte de los modelos objetivistas pre-cognitivos. Desde la interpretación operacionalista y acusadamente empiristá de la filosofía de la ciencia de la que partían muchos de esos modelos, era difícil dar cuenta de esas propiedades de lo mental. Esa dificultad se relacionaba con la tendencia a confundir los requisitos de justificación de los enunciados observacionales de la ciencia con los de justificación de sus conceptos teóricos. Es evidente que propiedades tales como la racionalidad y la intencionalidad *no* pueden reducirse a entidades directamente observables, sino que son inferencias netamente teóricas, que se emplean para describir formas peculiares de legalidad en la conducta de ciertos sistemas (en concreto, de aquellos a los que atribuimos conocimiento).

(3) Las consideraciones anteriores nos permiten establecer, de forma precisa, un aspecto esencial de los enfoques cognitivos clásicos en psicología (y dejando de lado, ahora, las perspectivas conexionistas más recientes): a diferencia de los planteamientos introspectivos, que empleaban enunciados observacionales intencionales, en primera persona de

singular, y de los conductistas, que trataron de reducir a enunciados extensionales tanto la base observacional como los conceptos teóricos en psicología, *los enfoques cognitivos clásicos (del Paradigma C-R, el Procesamiento de la Información e, incluso, precursores como Bartlett, Piaget y Vygotski) implican el empleo de conceptos intencionales para explicar observaciones establecidas en forma de enunciados extensionales.*

Un ejemplo puede aclarar fácilmente lo que queremos decir. Supongamos que queremos estudiar ciertos procesos de comprensión, y que, para ello, hacemos leer determinados textos a nuestros sujetos (que son «ellos», es decir, «terceras personas») y, transcurrido un tiempo, les pedimos que determinen si ciertos enunciados que les presentamos estaban o no explícitamente en los textos que leyeron. En cada uno de estos enunciados, determinamos si cada sujeto *acierta o falla* en la detección del enunciado como presentado explícitamente o no en el texto leído. Luego, para explicar el patrón de resultados obtenidos, construimos un modelo teórico que establece que, en función de sus *esquemas mentales* previos, y de la coherencia del texto con dichos esquemas, los sujetos establecen *modelos mentales* de los textos, en que infieren *proposiciones* no presentadas y eliminan proposiciones presentadas.

Los enunciados observacionales, que establecen que «los sujetos obtienen, por término medio (y con la desviación típica correspondiente) un número x de aciertos —o errores— en la detección como presentados de los enunciados z» son estrictamente extensionales. Los teóricos, que dicen que «los sujetos poseen esquemas/ infieren proposiciones, etc, que tienen esta o aquella forma y que se transforman, al ser procesados, de tal o cual manera», son *enunciados intencionales.* Porque los esquemas, las proposiciones mentales, los modelos mentales de los textos, los procesos de inferencia son objetos intencionales; es decir, sólo «son» si son sobre algo, si se refieren a contenidos, si «se llenan» de algún modo.

No es posible concebir una proposición que no sea *sobre* algo, un esquema que no sea *acerca de* algo, un modelo mental que no sea un modelo de algo. En el sentido en que aquí las empleamos, esas preposiciones, o locuciones preposicionales *{sobre, acerca de, de),* constituyen una marca lingüística de intencionalidad, mientras que no lo es la proposición «a» cuando decimos que «Alguien responde a algo», a no ser que se especifique que tal algo es un *contenido,* o el *significado* de un estímulo (lo que supone, en realidad, una intrusión subrepticia de intencionalidad en el caso de muchos modelos conductistas).

(4) Históricamente, la condición para el reingreso de objetos menta-

les (es decir, internos e intencionales) en psicología, después de su depuración conceptual por el conductismo y el objetivismo funcionalista de entreguerras, fue una doble condición:

(a) Tales objetos debían basarse y justificarse en enunciados extensionales sobre la conducta de las personas; en enunciados, por ejemplo, sobre *errores o latencias* en condiciones experimentales bien definidas y replicables. Así, un enunciado como «la latencia media de los sujetos del grupo A en la resolución de los problemas z es de 970 centisegundos» es absolutamente extensional.

(b) Por otra parte, los objetos teóricos de carácter intencional debían interrelacionarse de forma mecanicista (como hemos comentado extensamente en la sección segunda), es decir, por medio de procesos de cómputo determinados por las estructuras de tales objetos y~por reglas bien definidas.

Esta estrategia, que define la perspectiva de la psicología cognitiva de las computaciones sobre representaciones —e, implícitamente, del procesamiento de la información— desarrollada antes de las formulaciones conexionistas, deja en el aire algunas preguntas fundamentales a las que ya nos hemos referido en los dos capítulos anteriores. Por ejemplo las siguientes:

Si es sólo la *forma,* la *sintaxis,* de las estructuras intencionales de la mente la que determina el curso de los cómputos, entonces la intencionalidad (i.e. la posesión de contenidos) es una propiedad computacionalmente irrelevante de los objetos mentales. Ese supuesto implícito no sólo choca con el sentido común y las intuiciones de la psicología natural, sino que hace difícil explicar cómo han podido ser adaptativas las funciones mentales intencionales y epistémicas.

Pero, por otra parte, la idea de que ciertas propiedades extraformales, semánticas si se quiere, de los contenidos mentales puedan condicionar el funcionamiento mental no es nada fácil de compatibilizar con la hipótesis de que la mente es un sistema de cómputo. Ese es un aspecto de la crítica de Searle a la psicología cognitiva, a la que nos referíamos en el capítulo segundo, que no debe desdeñarse, porque apunta a uno de sus problemas esenciales (la solución de que todo el componente semántico de los contenidos mentales quizá se reduzca, en último término, a *dimensiones formales,* que propone Fodor en *Psychosemantics,* 1988, es, como hemos comentado, pura especulación).

Además, la solución «mecanicista sofisticada» de la psicología cognitiva dominante se encuentra en graves dificultades, que ya hemos visto,

para explicar la funcionalidad cognitiva de la conciencia y convierte la racionalidad en un supuesto de los algoritmos de cómputo, cuya génesis no es fácil de comprender. Así, aunque su vocabulario teórico muestre una semejanza aparente con el de la psicología natural, los supuestos tienden a ser claramente diferentes. Como ya señalaba en otra ocasión, el sujeto del que habla la psicología cognitiva no se identifica con el *sujeto personal* de la psicología natural y, como ha destacado Dennett (1978), las explicaciones cognitivas se sitúan en un nivel subpersonal:

> Si uno está de acuerdo con Fodor en que el objetivo de la Psicología Cognitiva es representar procesos psicológicamente reales que se dan en las personas, y puesto que la adscripción de creencias y deseos sólo está relacionada indirectamente con tales procesos, bien podemos decir que creencias y deseos no son objetos propios de estudio de la Psicología Cognitiva. Dicho de otro modo, las teorías cognitivas son o deben ser teorías del nivel sub-personal, en que desaparecen las creencias y los deseos, para ser reemplazados por representaciones de otros tipos y sobre otros temas (Dennett, *op. cit.,* p. 107).

Parece evidente que ninguna psicología científica puede conformarse con quedarse en el *nivel personal,* definido por la psicología natural de las actitudes proposicionales. Los enunciados fundamentales de esa psicología (es decir, aquellos que predican una mente intencional en primera persona de singular) son esencialmente *intratables* como enunciados observacionales básicos en psicología, aunque puedan tener un papel complementario importante.

La conceptualización de lo mental por la psicología natural es muy útil para la predicción y planificación de conductas en situaciones corrientes de interacción, pero ofrece dificultades graves para la teorización psicológica: se fundamenta y justifica en la versátil fenomenología de la conciencia, que discrimina acerca de relaciones con objetos mentales más que acerca de la mente que los construye y transforma. Además, esa conceptualización permite inferencias limitadas e idiográficas y no busca leyes causales sino predicciones de conductas a corto plazo. Se maneja con supuestos teleológicos difíciles de incluir en las relaciones causales de la ciencia y articula formulaciones *narrativas* más que relaciones *paradigmáticas* (Bruner, 1986) acerca de lo mental. Así, la psicología natural parece estar diseñada para *comprender* la mente, pero no para *explicarla* (siguiendo la distinción que dio lugar a la conocida polémica entre Ebbinghaus y Dilthey). El problema de cómo construir una psico-

UNIVERSIDAD DE ANTIOQUIA
BIBLIOTECA CENTRAL

logía explicativa y objetiva, sirviéndose de las herramientas teóricas proporcionadas por los enunciados intencionales, remite al origen de la psicología cognitiva como ciencia objetiva de la mente, al que nos referiremos, de nuevo, al final de nuestras reflexiones.

Capítulo 9

EL PENSAMIENTO Y EL LENGUAJE COMO FUNCIONES MENTALES. ALGUNAS REFLEXIONES SOBRE INTENCIONALIDAD, INTENSIONALIDAD Y FORMA

9.1. Pensamiento y lenguaje: algunos enfoques de la «psicología popular»

El ejercicio que hemos hecho en los capítulos anteriores, consistente en dibujar el paisaje conceptual que empleamos en nuestras relaciones cotidianas como psicólogos naturales, puede proporcionarnos un marco de referencia muy útil para comprender el origen y la naturaleza de la psicología cognitiva. La razón principal es que no podemos dejar de ser psicólogos naturales cuando hacemos una ciencia objetiva de la mente. Aun cuando adoptemos una actitud extremadamente reduccionista y objetivista, una visión «extensionalista» de lo psicológico, las categorías mentalistas que forman parte de nuestro sistema cognitivo, en su calidad de mecanismos de cómputo y predicción psicosocial, se inmiscuyen constantemente en nuestro pensamiento y destacan su presencia, unas veces como blancos a abatir, otras como conceptos que deben ser elaborados y refinados, algunas como presuposiciones implícitas que condicionan nuestras formas de entender los fenómenos psicológicos. Están ahí, de forma inevitable, a favor o en contra de nuestro pensamiento y juegan, con frecuencia, un papel ambiguo de obstáculos y ayudas, al mismo tiempo, al que ya nos hemos referido.

Por todo ello, puede resultar esclarecedor el análisis de algunos conceptos científicos de la mente, empleando las mismas herramientas conceptuales que hemos ido elaborando para comprender la psicología natural. En este capítulo nos referiremos principalmente a *ciertas funciones mentales:* las relacionadas con el pensamiento y el lenguaje, y no a otras

funciones, como las relacionadas con la percepción y la memoria, a pesar de la indudable importancia cognitiva que éstas pueden tener.

Las funciones de pensamiento y lenguaje presentan peculiaridades que permiten afirmar que constituyen una clase definida, incluida dentro de la más general de las funciones mentales. Mantienen, entre sí, relaciones complejas, profundas, y no comprendidas aún en su totalidad, que hacen muy exacta la afirmación reciente de Asensio (1990): «El Pensamiento y el Lenguaje parecen una de esas parejas que se separan por no poder vivir juntos y que, al cabo del tiempo, se vuelven a juntar porque no pueden vivir separados (y en cada separación se reproducen las discusiones sobre las gananciales, la casa, los hijos)» (p. 9).

La tendencia más primitiva de la «psicología popular», en la conceptualización de las relaciones entre pensamiento y lenguaje parece implicar, frecuentemente, un reflejo de la intuición de la profundidad de estas relaciones. Esa tendencia se expresa en la indiferenciación entre pensamiento y lenguaje (o más bien entre pensamiento y habla) que se encuentra en muchas culturas primitivas y en concepciones infantiles de lo mental.

Pueden servir como ejemplo, de la indiferenciación mencionada, algunas ideas que los nativos de las Islas Trobriand, cercanas a Nueva Guinea, tienen acerca de la localización de lo mental: «La mente, *nanola,* término con el cual se describen la inteligencia, el poder de discriminación, la capacidad de aprender fórmulas mágicas y todos los tipos de habilidades no manuales, así como las cualidades morales —dice Malinowsky— se encuentra en algún punto de la laringe. Los nativos apuntarán siempre a los órganos del habla, donde se encuentra la *nanola.* El hombre incapaz de hablar /../ queda identificado por el nombre (*tonagowa)* y el tratamiento con todos los que presentan deficiencias mentales» (1922, p. 408). Una concepción semejante a la que parecen tener los gahaku-gamas, de Nueva Guinea, para los que «saber» y «pensar» se identifican con «oír» (*gelenowe*) (Read, 1955, citado en Hallpike, 1981).

De forma parecida, los griegos de tiempos de Homero tendían a identificar el pensamiento con el aliento, y el acto de «pensar» con el de «hablar». El aliento se concebía como la inteligencia de la palabra. Onians (1954) afirma que «la creencia de que los pensamientos son palabras, y las palabras aliento /.../ lleva a la conclusión (entre los griegos) de que los órganos de respiración, los pulmones, son los órganos de la mente» (p. 68).

Estas ideas son, por otra parte, muy similares a las descritas por Piaget en su estudio sobre *La representación del mundo en el niño* (1926, vid. edición española de 1984) en que investiga la noción que tienen los niños occidentales sobre el pensamiento:

MONT (7; 0). (Pregunta Piaget)

> ¿Sabes lo que es pensar? —Sí. —Piensa en tu casa, ¿quieres? —Sí. —¿Con qué piensas? —Con la boca. —¿Puedes pensar con la boca cerrada? —No. —¿Y con los ojos cerrados? —Sí. ¿Y con las orejas tapadas? —Sí. —Cierra la boca y piensa en tu casa. ¿Piensas? —Sí. —¿Con qué has pensado? —Con la boca (p. 43).

Es evidente que la fusión de pensamiento y los productos de la boca, la laringe, los pulmones, la identificación del pensamiento con la voz, la audición o la respiración, son expresiones de modelos de la mente sincréticos en que los fenómenos físicos no se diferencian, con claridad, de las propiedades mentales de ciertos objetos; manifestaciones de un pensamiento que, al tiempo que atribuye mente a los fenómenos naturales, se fundamenta en una interpretación material y sustancial de lo mental (de la que proviene la tendencia a «materializar» en la respiración, la laringe, etc, el pensamiento).

Sin embargo, también se expresan, en esas manifestaciones de la «psicología popular» primitiva intuiciones acerca de las densas interrelaciones que existen entre las funciones que se señalan, en todos los lenguajes naturales, en los verbos de Pensamiento y Lenguaje: pensar, inferir, decir, comprender, razonar, etc. Antes de revisar, en el capítulo próximo, algunas de las concepciones desarrolladas por la psicología científica sobre estas funciones, conviene que nos detengamos en un análisis, lo más escueto posible, de algunas características comunes al pensamiento y el lenguaje y, al mismo tiempo, específicas de estas funciones, desde la perspectiva, que hemos utilizado hasta ahora, de la psicología intencionalista de las actitudes proposicionales.

9.2. **Pensamiento, lenguaje e intensionalidad**

Decíamos, en otro momento, que la propiedad psicológica esencial de las funciones mentales, la *intencionalidad* (es decir, la propiedad de «ser acerca de algo»), se refleja en una propiedad lógica de algunos

enunciados con verbos mentales: la *intensionalidad.* Y definíamos las características de los enunciados intensionales, resumiéndolas en tres aspectos:

(1) Cuando un enunciado es intensional (Pedro cree que el Mago de Hoz es ambidextro; María dice que la Bruja Piruja es malvada), la verdad del enunciado completo (Pedro cree que X, María dice que Y) *no implica:*

(a) ni la verdad de las proposiciones sometidas a la cláusula intencional (El Mago de Hoz es ambidextro, la Bruja Piruja es malvada),

(b) ni la existencia de los sujetos de tales proposiciones (El Mago de Hoz, la Bruja Piruja),

(c) ni la posibilidad de sustituir tales sujetos por otros con el mismo referente pero distinto significado (El Mago de Hoz por «el Mago que aparece en el libro de Frank Baum titulado *The Wonderful Wizard of Oz*» o «El Mago que aparece en la película de Victor Fleming de 1939»),

Ahora bien, esta propiedad de intensionalidad, que es, por así decirlo, la *marca lógica* de la intencionalidad, se produce, ante todo, en los enunciados con verbos de Pensamiento y Lenguaje. En los de percepción no se da (Searle, 1983).

Así, si es verdad que «alguien percibe X», ello presupone la existencia de X, y si X no existe, lo que hace ese «alguien» no es percibir sino alucinar, de manera que sería semánticamente anómalo decir que «alguien percibe X», cuando no existe X. Además, del X que se percibe no podemos predicar, en rigor, la propiedad de ser verdadero o falso, ya que las proposiciones no se perciben; se perciben objetos; la percepción, en su sentido preciso, es una función anterior a la predicación de propiedades en los objetos (quizá pueda decirse que «asigna» propiedades, de forma implícita, pero no que las «predica»). Finalmente, el criterio de no-sustitución carece, por todo lo anterior, de sentido cuando se aplica a enunciados de percepción, ya que éstos se refieren precisamente a los procesos mentales que configuran referentes y tales procesos son, en cierto modo, previos a (aunque no independientes de) aquellos que asignan significados.

Los verbos relacionados con la memoria parecen tener una intensionalidad parcial (pues tampoco resulta adecuada semánticamente, en sentido estricto, la afirmación de que «alguien recuerda X» cuando X no fue el caso, mientras que sí lo son las proposiciones de que «alguien piensa X» o «dice X», cuando X no es el caso). En suma: los enunciados sobre pensamiento y lenguaje son los que poseen, de forma clara e indiscutible, una marca lógica de intencionalidad.

Ello se relaciona con dos aspectos estrechamente relacionados entre sí que tienen las funciones mentales de tipo P-L: (1) son actividades simbólicas por antonomasia (y ya hemos apuntado, en otro momento, la relación existente entre la actividad simbólica y la propiedad lógica de intensionalidad) y (2) producen objetos intencionales que desbordan, por completo, el mundo limitado de los referentes creados por la percepción.

9.3. **La intuición de la forma en el pensamiento** y **el lenguaje**

El pensamiento puede producir, y el lenguaje expresar, objetos y relaciones que poseen propiedades epistémicas peculiares. Al mismo tiempo, van «más allá de la información dada», a través de los mecanismos perceptivos, y se acompañan, en ocasiones, de una impresión neta de validez o necesidad (por ejemplo, de validez y necesidad lógica) que no alcanzan nunca los objetos y relaciones, de carácter fenoménico, que la percepción establece. Se puede expresar, de forma muy sintética, esta idea, diciendo que pensamiento y lenguaje, y sólo ellos, son —pueden ser— *portadores de razón.* Y esta propiedad, que no tienen en sí mismos los objetos contingentes de la percepción y la memoria, ha atraído siempre a la reflexión sobre el conocimiento y ha llegado a adquirir una importancia esencial en algunas ontologías, como la de Platón.

La razón es una facultad de las personas, pero también una propiedad de algunas de las relaciones que se piensan y se dicen; a saber, la propiedad de tales relaciones de poseer una validez de carácter universal (ver el apartado 2. del capítulo segundo) y extramental. Una validez que aparenta ser distinta y superior, por su carácter necesario, a la que tienen los objetos conocidos por la percepción y la memoria episódica.

Cualquier psicología del conocimiento tiene que dar cuenta de esa propiedad, que históricamente ha demostrado ser difícil de reducir a un efecto de mecanismos puramente asociativos y a enfoques epistemológicos limitadamente empiristas. Hay objetos, que se crean mediante el pensamiento —tales como un triángulo, la bondad, la justicia, y la inferencia de que si «todos los A son B», y «X es A», entonces «X es B»— que no se imponen por su evidencia perceptiva sino por una exigencia lógica inmune a las volubles contingencias de las cosas.

Ello plantea un difícil desafío a la psicología del conocimiento: el de explicar cómo un sistema que se relaciona con el medio a través de sus acciones y percepciones puede llegar a establecer, primero, y aplicar a «lo real» esos objetos y relaciones. Si todas las «entradas» con que cuenta el sistema cognitivo provienen, en último término, de los procesos de transducción sensorial y de configuración de un mundo completamente contingente de objetos y relaciones perceptivas, ¿cómo es posible que el propio sistema cognitivo acceda a objetos que no son perceptibles (como un espacio de más de tres dimensiones), o a relaciones que no son contingentes (como las que establecen las inferencias matemáticas)?

Esta «pregunta de Platón» ha recibido respuestas muy diversas en dos mil quinientos años de especulación sobre el conocimiento. Parece apuntar a la existencia de objetos que no se originan en la percepción ni en la memoria de los episodios contingentes del mundo. Pero esa ilusión platónica del origen extramental de los objetos y las relaciones necesarias no puede satisfacer plenamente al psicólogo. Toda la psicología genética de Piaget, por ejemplo, es un intento de dar una respuesta a esta cuestión fundamental, a la que las concepciones estrechamente «psicologistas» del conocimiento nunca han dado una solución convincente.

La propiedad que posee el pensamiento humano de producir objetos intencionales de validez extramental (y cuyo origen no reside en la semántica de la percepción) tiene una relación muy clara con otra característica a la que debemos referirnos: consiste en el hecho de que el pensamiento y el lenguaje poseen una *forma* en un sentido diferente a cómo la poseen los objetos perceptivos. Una forma que no depende de la naturaleza «holística», «semánticamente densa», analógica, de las representaciones, sino del hecho de que éstas se componen de constituyentes con relaciones bien definidas que corresponden —o pueden corresponder— a *reglas.*

La idea de que existen unidades diferenciadas de representación, de carácter analítico, divisibles en partes claramente diferenciadas, obedientes a reglas, es aplicable, sobre todo, al tipo de unidades de representación que se manejan en las funciones de pensamiento y lenguaje o, por decirlo mejor, que definen a estas funciones como tales. No tienen sentido los conceptos de pensamiento y lenguaje si no hacen referencia a ciertas estructuras que, a la vez que representan intencionalmente, poseen una forma y se transforman en función de reglas formales sobre estructuras formales.

Los actos de pensar, decir y comprender se acompañan, con frecuencia, de la intuición de corresponder —o no— a ciertos sistemas formales, de carácter canónico: en ocasiones intuimos que una inferencia no es válida porque sus proposiciones no se encadenan de forma adecuada; a veces, sentimos que una oración «no está bien construida», porque no se corresponde con nuestras intuiciones de gramaticalidad en nuestro lenguaje natural. Estas son intuiciones formales. En el sistema lingüístico de Chomsky (1957, 1965), por ejemplo, esas intuiciones de gramaticalidad constituyen la «base observacional» en que se fundamentan las construcciones de la gramática. Y, en todas las lógicas, las intuiciones de necesidad inferencial, han jugado un papel decisivo en la elaboración de cálculos.

Esto no sucede con los objetos de la percepción —al menos,, no sucede en el mismo sentido—, de los que no podemos decir que estén bien o mal construidos en función de relaciones precisas entre partes atómicas y descomponibles,.sino «bien o mal formados» en virtud de propiedades «holísticas», que se deshacen al ser analizadas, como las que definieron los psicólogos de la Gestalt.

De los resultados del pensamiento y el lenguaje sí podemos decir que están bien construidos o mal construidos. Y ello significa que, implícitamente, la psicología natural parte del supuesto de que existen *modelos formales,* que se ven mejor o peor reflejados en tales actos. Naturalmente, esos modelos formales son los que elaboran, de forma explícita, los lógicos y los gramáticos. Así, los psicólogos que tratan con el pensamiento y el lenguaje se ven enfrentados a una necesidad que no tienen, en igual medida, otros psicólogos: la de definir, de forma precisa, las relaciones entre sus modelos de la mente y esos modelos canónicos, formales, que desarrollan los lógicos y los lingüistas.

¿Son esos modelos, en rigor, modelos de la mente (como, por ejemplo, pretende frecuentemente Chomsky de sus formalizaciones lingüísticas) o descripciones estructurales de sus productos que no tienen relación con las funciones mentales en sí mismas?; ¿hasta qué punto es lógico, y en qué condiciones lo es, el pensamiento humano, y reductible a formas gramaticales, inequívocamente regladas, el lenguaje? ¿Hasta dónde puede reducirse el problema de la forma gramatical del lenguaje al de la forma lógica del pensamiento?; ¿qué relaciones deben darse entre la psicología empírica y objetiva del conocimiento y esas otras ciencias —lógica y lingüística— que definen formas canónicas de inferencia y gramaticalidad?

9.4. La intuición de la forma y la psicología cognitiva

La «intuición de determinación por la forma», que se manifiesta con frecuencia en el ejercicio de las funciones de pensamiento y lenguaje, ha tenido una enorme importancia en los paradigmas C-R y de procesamiento de la información de la psicología cognitiva. La razón es fácil de comprender: esa intuición ha servido de fundamento para el desarrollo de los modelos computacionales de la mente, que parten de la hipótesis de que la forma de las representaciones determina las reglas que se siguen en los procesos de cómputo.

Cuando se utiliza en este contexto, el concepto de «forma» no tiene el sentido de «emergente» de propiedades holísticas de objetos analógicos que tiene en la-psicología de la Gestalt, sino que apunta a mecanismos propiamente constructivos, semejantes a los que se ponen en juego cuando se construyen oraciones gramaticales de un lenguaje, estando formadas dichas oraciones por cláusulas estrictamente delimitadas, frases de fronteras netas, palabras o elementos léxicos que definen un cierto plano «atómico» del lenguaje representacional.

Por ello, se ha dicho, a veces, que los enfoques dominantes en psicología cognitiva incorporan un «modo lingüístico de explicación» (Carello, Turvey, Kugler y Shaw, 1982): definen modelos que consisten en algoritmos que aplican procesos discretos de cómputo a representaciones analíticas, semejantes a oraciones o proposiciones —y no, por ejemplo, semejantes a objetos visuales—, guiándose por reglas bien definidas y determinadas por la forma de las representaciones.

Más en general, podríamos decir que, en estos enfoques, sucede que el pensamiento y el lenguaje se convierten en *paradigmas* de lo mental. Esta caracterización permite establecer una diferencia neta entre los modelos computacionales de la mente y los que predominaron en la fase pre-cognitiva de la psicología: en éstos, las funciones de pensamiento y lenguaje tendían a asimilarse a modelos teóricos proporcionados por otros tipos de funciones, como la percepción (Gestalt), el aprendizaje o los procesos de condicionamiento (conductismo) o los mecanismos asociativos de la memoria episódica (funcionalismo objetivista de los años cuarenta y cincuenta). Por el contrario, los modelos cognitivos emplean frecuentemente metáforas del pensamiento y el lenguaje para explicar otras funciones: la percepción es inferencial (vid. García Albea, 1986), el aprendizaje implica descubrimiento (Pozo, 1987), la memoria reconstruye inferencialmente sus objetos y se guía por estrategias de razonamiento (Lindsay y Norman, 1972), todos los sistemas de representación pueden reducirse a símbolos atómicos y analíticos (Pylyshyn, 1988),etc.

Los modelos «lógico-lingüísticos» de explicación, difícilmente evitables cuando la mente se concibe desde una perspectiva computacional, han dado lugar —especialmente en la primera fase de desarrollo de esa perspectiva— a una consecuencia natural: el *logicismo,* basado en el metapostulado terminal ya comentado, según el cual «Las representaciones y/o procesos mentales humanos son isomorfos respecto del sistema de reglas formales lógico o matemático» (De Vega, 1981, p. 347).

Este postulado, fuertemente interiorizado e influyente en muchos modelos cognitivos de la formación de conceptos (Bruner, Goodnow y Austin, 1960; Bourne, 1968), la representación del conocimiento (Anderson, 1976, 1978; Pylyshyn, 1986), el desarrollo intelectual (Piaget, 1949, 1953), el razonamiento (Henle, 1962) y las actividades de predicción (Edwards, 1968; Peterson y Beach, 1967) puede considerarse paralelo a un metapostulado *gramaticalista* en psicolingüística, que establecería que «los procesos y representaciones mentales en el procesamiento psicológico del lenguaje son isomórficos con las reglas de transformación definidas por la gramática generativo-transformacional», y que se defendía explícitamente en las investigaciones inaugurales de la psicolingüística realizadas por investigadores como Miller y McKean (1964), Savin y Perchonock (1965)), Mehler (1963), Slobin (1966), etc (vid. revisión de Greene, 1972).

Las elaboraciones teóricas y estudios empíricos, encaminados a definir sistemas naturales de cómputo capaces de producir objetos con propiedades lógicas y gramaticales han sido, desde el principio de la psicología cognitiva, muy influyentes por sus implicaciones teóricas para el estudio de las funciones de Pensamiento y Lenguaje, pero también por su significación para la comprensión de otras funciones. Todos estos estudios se basaban en una hipótesis generativa, según la cual es posible formular procedimientos efectivos que modelarían las inferencias humanas (lógicas o no) y producirían las oraciones gramaticales de los lenguajes naturales.

Así, la psicología cognitiva del procesamiento de la información ha tendido a comprometerse con propensiones formalistas, logicistas y gramaticalistas que —y esto es lo importante— se derivaban aparentemente de sus hipótesis más fundamentales y, más en concreto, de la metáfora del ordenador, a la que nos hemos referido más extensamente en la sección segunda de este libro.

El pensamiento y el lenguaje son, como ya comentábamos en otro

momento, las funciones más exigentes en cuanto a las propiedades semánticas que deben poseer los objetos a que se atribuyen. Son aquellas que más nos hemos resistido a aplicar, los miembros perspicaces pero presumidos de esta especie nuestra, a los de otras especies (vid., sin embargo, Griffin, 1984, y Premack y Premack, 1983). Poseen, en el más alto grado, esa marca de intencionalidad que es la intensionalidad lógica; configuran objetos que parece reflejar la presencia invisible, pero evidente, de modelos formales y canónicos, con cierta consistencia y validez extramental. Son, sin duda, las funciones mentales que aparentan ser fenomenológicamente más complejas (aunque no tengan por qué ser las computacionalmente más complejas) y están tan entrelazadas con los productos de la sociedad y la cultura que parecen especialmente difíciles de explicar por una ciencia causal y objetiva de la mente.

¿Cómo ha explicado la psicología las funciones de pensamiento y lenguaje? ¿Cómo se ha enfrentado la psicología empírica del conocimiento a esas difíciles propiedades de complejidad, carácter simbólico, dependencia socio-cultural e intencionalidad peculiar de los objetos de las funciones de pensamiento y lenguaje? Dedicaremos a estas cuestiones el capítulo final de nuestra reflexión.

La historia del intento de dar una explicación objetiva, científica, de qué significan realmente los verbos mentales más paradigmáticos, los relacionados con pensar, comprender y decir, representa un momento apasionante de la ciencia: aquel en que el conocimiento objetivo se enfrenta a los procesos por los que la subjetividad se define de forma más neta y en que la misteriosa propiedad de intencionalidad alcanza su expresión más alta.

Capítulo 10

CUATRO ENFOQUES PSICOLOGICOS DE LAS FUNCIONES MENTALES

10.1. Introspección, objetividad y sistemas de cómputo

Las reflexiones que hemos ido haciendo, a lo largo de este libro, pueden proporcionarnos instrumentos conceptuales útiles para examinar, desde una perspectiva nueva, algunos episodios importantes en la historia de la psicología científica. La pretensión de este capítulo es la de presentar un boceto general de esa reinterpretación de la historia de la ciencia de la mente, y no la de ofrecer el cuadro acabado de un análisis histórico que desbordaría con mucho las dimensiones razonables de esta obra.

Comentábamos, en un capítulo anterior, que en realidad los psicólogos siempre nos hemos ocupado de los objetos con mente y de las funciones mentales. Decíamos que nuestro campo de interés puede definirse de forma muy breve: Fm. Sin embargo, el modo de entender esas funciones ha variado mucho a lo largo de la historia. En el transcurso de ella, sobresalen cuatro episodios fundamentales, a los que ya nos hemos referido en el capítulo octavo. Son los siguientes:

(1) El primero, definido por la psicología introspectiva de finales del siglo XIX y comienzos del veinte, implicó aceptar enunciados con verbos mentales, o con entidades intencionales, tanto en los supuestos observacionales como en el vocabulario teórico de la psicología. Eran aquellos los tiempos en que aún existía la ilusión de poder acceder a un conocimiento científico de la mente a través de la autoobservación o de la percepción interna. Por medio de la introspección, en una palabra.

(2) El segundo episodio es el que se refleja principalmente en el conductismo y en algunas tendencias objetivistas de la psicología sovié-

tica. Su marco temporal se sitúa principalmente en las décadas transcurridas entre los años de la primera guerra mundial y los sesenta. El desaliento producido por el descrédito de la psicología introspectiva se tradujo en un intento de construir algo así como una «psicología sin mente». Una psicología en que los verbos y entidades mentales (las imágenes, los planes, los pensamientos, las intenciones, etc) se consideraban bien como términos y entes directamente elimihables del vocabulario riguroso de la psicología, o bien como términos a traducir a un lenguaje extensional preciso: el de los reflejos, los estímulos y las respuestas.

La pretensión principal de estos enfoques era; así, la de realizar una tarea de depuración conceptual que eliminase las dos propiedades esenciales (intencionalidad e interioridad) de lo mental. El supuesto básico era el de que tanto los enunciados de observación como las afirmaciones teóricas de la psicología deben realizarse en términos estrictamente extensionales.

(3) El tercer episodio corresponde ya a la jetapa cognitiva y a la segunda mitad de nuestro siglo. Parte de la metáfora —o consideración literal— de la mente como un sistema de cómpilo que trata con **representaciones,** es decir, con objetos intencionales **yj** simbólicos que reciben diferentes nombres: esquemas, proposiciones, conceptos, etc.

Implica un cierto compromiso entre las dos posiciones anteriores, pero un compromiso que es sólo aparente: en el aspecto teórico, se sirve, como la psicología introspectiva, de un vocabulario intencional. En el empírico, se basa, como el conductismo, en un vocabulario establecido generalmente en términos extensionales, y que incluye elementos tales como las latencias y errores de los sujetos ante tareas bien definidas.

Sin embargo, esa posición no es una mezcla ecléctica de las anteriores. El compromiso es, ciertamente, sólo aparente: ni el vocabulario empírico de la psicología cognitiva tiene el mismo significado que el del conductismo, ni el teórico el mismo que tenían los términos intensionales en las psicologías mentalistas de corte introspectiva. La nota de mecanicismo, a la que ya nos hemos referido extensamente, ayuda a entender hasta qué punto existe una profunda diferencia entre los conceptos teóricos introspectivos (ajenos del todo a la noción de cómputo, y muy enraizados en la fenomenología consciente) y los cognitivos. Por otra parte, los enunciados empíricos extensionales de la psicología cognitiva clásica —a diferencia de los del conductismo— sólo encuentran significado cuando se insertan en un rico complejo de presupuestos teóricos y términos explicativos de naturaleza mentalista.

(4) Finalmente, el cuarto episodio es mucho más reciente, y sólo puede entenderse cuando se mira desde el marco histórico y conceptual de los dos anteriores:, se trata del episodio del conexionismo, al que ya nos referíamos en el capítulo cuarto de este libro. Representa, desde la perspectiva de análisis que estamos empleando un cambio fundamental: la vuelta a un vocabulario teórico estrictamente extensional (es decir, no compuesto de representaciones simbólicas semánticamente transparentes, sino de fuerzas de conexión entre unidades subsimbólicas, niveles y patrones de activación, umbrales de in-put y out-put de tales unidades), pero que admite que los objetos intencionales son resultantes del funcionamiento masivo y paralelo de unidades, configuradas en redes organizadas y definidas en términos puramente cuantitativos.

Del mismo modo que no puede entenderse el origen de la psicología cognitiva clásica si no se toma en consideración la depuración metodológica previa realizada por el conductismo, al prescindir de la introspección y de enunciados observacionales mentalistas, tampoco puede entenderse el desarrollo reciente de los modelos conexionistas, si no se recuerda el hecho de que éstos se plantean, como pregunta fundamental, una pregunta acerca de la mente **computacional** (y no acerca de la mente **fenoménica)** que se realizó, por vez primera, desde los modelos cognitivos clásicos del modelo C-R y la Psicología del Procesamiento de la Información.

El análisis anterior nos permite entrever un aspecto importante en el desarrollo de los episodios a que haremos referencia a lo largo de este capítulo: con excepción de la «revolución conductista», la transición entre el propio conductismo y la psicología cognitiva clásica, por una parte, y entre ésta y el conexionismo, por otra, implica, en aspectos importantes, una transición en que se conservan aspectos importantes, en cada caso, de la posición anterior.

La doble tendencia a prescindir (aunque no necesariamente del todo) de enunciados mentalistas introspectivos para definir la base observacional de datos de la psicología, y a considerar que la mente es, en una u otra forma, un sistema de cómputo, define características esenciales de la psicología contemporánea. Conocer objetivamente esa **mente computacional,** de la que venimos hablando, es el objetivo esencial que nos hemos propuesto los psicólogos de la segunda mitad de nuestro siglo.

10.2. Funciones mentales y psicología introspectiva: consideraciones históricas

Decíamos, pocas páginas más atrás, que la psicología natural parte de la intuición de que los enunciados mentales se justifican, de forma esencial y prototípica, cuando se hacen en primera persona de singular, y añadíamos que la primera persona de singular es el núcleo básico, la «residencia natural» de los enunciados psicológicos.

Esta intuición se basa probablemente en el hecho de que la conciencia reflexiva es, entre otras cosas, un mecanismo metarepresentacional destinado a la previsión de las conductas propias y ajenas, que se sirve del plano mental que tiene el sujeto de sí mismo, para fundamentar sus predicciones sobre otros objetos que son considerados también como sujetos. Así, no es extraño que los primeros intentos de hacer psicología científica trataran de hacer compatible la objetividad necesaria de la ciencia con la poderosa intuición de que la primera persona (y, por consiguiente, la introspección consciente) es la raíz que fundamenta la predicación de funciones mentales. Era ésta uña tarea difícil, y no creo exagerar si digo que se saldó históricamente con un magnífico fracaso, y con la eliminación objetivista de los conceptos mentales en buena parte de la psicología.

Con el vocabulario empleado hasta ahora, podemos decir que la introspección es el proceso por el cual se pasa, desde los enunciados mentales en primera persona, a otros enunciados que tratan de aclarar en qué consisten las propias funciones mentales: se va de «yo pienso (o percibo, recuerdo, etc) X» a «mi pensar (recordar, etc) consiste en...». **Este salto presupone que en la conciencia de la relación con objetos que la mente elabora, se incluye una conciencia de los procesos o estructuras que permiten esa elaboración.**

Planteadas así las cosas, el problema con el que siempre se encontró la psicología introspectiva fue el de establecer las condiciones y justificaciones que debían darse para pasar del primer tipo de enunciados («Yo Fm X»), a los que llamaremos «enunciados 1», al segundo («Mi Fm consiste en Y», o «enunciados 2»). Y, en contra de la imagen simplista y poco articulada de la psicología introspectiva que se ha ofrecido tantas veces, hay que decir que, con relación a ese problema central (el paso de los enunciados 1 a los 2), las posiciones de los psicólogos introspeccionistas fueron muy diversas y, en muchos casos, demostraban una sensibilidad muy acusada con relación a los problemas metodológicos de la introspección.

Resumiendo mucho, podemos establecer, al menos, dos grandes enfoques de la introspección:

(1) Un enfoque liberal, más Seguro de la posibilidad de pasar de los enunciados 1 a los 2, y más ingenuamente confiado en los aspectos fenoménicos de lo mental.

(2) Un enfoque restrictivo, que establecía límites muy estrictos a las condiciones experimentales, los valores del conocimiento fenoménico y las posibilidades o ámbitos de estudio de la introspección.

El primer enfoque se desarrolló, sobre todo, a partir de la tradición filosófica del empirismo inglés y de la Escuela escocesa del sentido común. El segundo, elaborado sobre todo por Wundt, era tributario de tradiciones germánicas de pensamiento, ligadas a ideas de Leibniz, Kant y Herbart, entre otros. A la vez que atribuía mente a todo lo natural, Leibniz negó la identidad mente-conciencia en que se basó tanto el pensamiento cartesiano como el empirismo británico. Esta identidad fue cuestionada también por Herbart. Por su parte, Kant fue muy severo con relación a las posibilidades científicas de la introspección que, para él, sólo podía acceder a un conocimiento limitado y casual del «yo fenoménico», siéndole inaccesible el verdadero fundamento de la vida mental, el sujeto que percibe, categoriza, juzga y razona (este es uno de los aspectos en que la posición de Kant se parece bastante a la de los psicólogos cognitivos actuales).

Este marco conceptual, definido por posiciones en que se cuestionaba o limitaba mucho el papel de la introspección, influyó claramente en los enfoques de Wundt. Este se refirió a las limitaciones metodológicas de la introspección con ocasión de dos debates: el primero, a raíz de la creación del laboratorio de Psicología experimental y de la revista *Philosophische Studien,* le enfrentó a algunos filósofos idealistas, que ponían en duda la posibilidad de una psicología experimental. El segundo, le opuso a los psicólogos de la escuela de Würzburgo. En esencia, la posición de Wundt se basó en la distinción, a la que ya hemos aludido entre «autoobservación» *(Selbstbeobatchung)* y «percepción interna» (*innere Wahnehmung).* La primera era susceptible de las críticas antiintrospectivas de Lange y Comte (no establecería una distinción suficiente entre sujeto y objeto de observación, y alteraría, hasta hacerlos irreconocibles, los fenómenos observados). Por el contrario, la percepción interna de los fenómenos de la mente era, para Wundt, una realidad incuestionable, que, empleada con las debidas garantías, sí podía constituir una base (no la única) de la psicología científica.

¿Y cuáles eran esas garantías? Para Wundt, la única forma de hacer psicología científica con la introspección consistía en manipular las condiciones de la «percepción interna» de tal forma que se acercasen lo más posible a las de la externa, es decir, a las que caracterizan a los enunciados observacionales de las ciencias de la naturaleza. Ello implicaba tres operaciones:

(1) Reducir al máximo el tiempo entre la percepción original y el informe acerca de ella.

(2) Asegurar las condiciones de replicabilidad y simetría.

(3) Limitar drásticamente las funciones estudiables por introspección.

Estas exigencias limitaban prácticamente la utilidad de la introspección al estudio de la percepción, en que podía establecerse una correspondencia relativamente clara, y muy directa, entre los enunciados intencionales y las condiciones estimulares definidas extensionalmente. De este modo, la introspección no era útil para estudiar las funciones mentales más complejas de pensamiento y lenguaje.

De hecho, y como ha destacado Danziger (1980), la mayor parte de los estudios del laboratorio de Wundt trataron de temas relacionados con la sensación e implicaron el empleo de medidas objetivas, cuyos resultados se establecían en términos rigurosamente extensionales, de tiempos de reacción:

> De cerca de 180 estudios experimentales publicados entre 1883 y 1903 en los veinte volúmenes de los *Philosophische Studien* —dice Danziger— sólo hay cuatro en que se usan datos cualitativos introspectivos de forma semejante a como se utilizaron en la práctica en la década siguiente. Por otra parte, el laboratorio de Wundt produjo una gran cantidad de estudios cuyos datos eran estrictamente «conductuales», la mayor parte establecidos en forma de diversas clases de medidas de tiempos de reacción. Lo que era «mentalista» en aquellos estudios era la interpretación teórica de los resultados, y no los datos en sí mismos (p. 248).

Así, Wundt fue, de hecho y con independencia de sus pretensiones teóricas, un verdadero precursor de los enfoques cognitivos, caracterizados por el empleo de conceptos intencionales para explicar datos extensionales.

Naturalmente, los estudios sobre el lenguaje y sobre «los procesos de pensamiento en sí mismos» (que, para Wundt, sólo podían estudiarse a través del lenguaje) no eran accesibles a la metodología introspectiva y

estaban fuera del alcance de la Psicología experimental. En esto, Wundt mantenía una posición muy diferente a la de los psicólogos cognitivos de hoy. El estudio del lenguaje debía situarse en el ámbito de la Psicología de los Pueblos, más cercana por su metodología a las Ciencias del Espíritu (*Geisteswissenschaften*) que a las de la Naturaleza (*Naturwissenschaften*). «De acuerdo pues con las Ciencias del espíritu —dice Miralies— la conciencia adulta desarrollada no puede ser abordada mediante el análisis, ya que sus componentes no son materiales. El paralelismo, en este caso, no se daría entre procesos psíquicos y fisiológicos, sino que el lugar de éstos lo toman los eventos históricos y colectivos que no se pueden reducir a una suma de elementos últimos. En la medida en que la cultura es un resultado del espíritu de la sociedad, los métodos adecuados para su estudio son los comparados e históricos» (1986, p. 60). Para Wundt, como para Vygotski, era necesario «salir fuera» de los límites de lo subjetivo y remitirse a las formas de la vida social para dar cuenta de la génesis y naturaleza de las funciones superiores de Pensamiento y Lenguaje.

Wundt preveía que la aplicación de la metodología introspectivo-experimental a las funciones superiores iba a ser una fuente de problemas, como efectivamente sucedió, cuando los psicólogos de la escuela de Würzburgo, empleando una forma «liberal» de introspección, a la que se dio el nombre de «introspección sistemática», trataron de enfrentarse con ella a las funciones de pensamiento.

Como ha mostrado Danziger (1980), en la primera década de nuestro siglo, se impusieron formas de introspección que respondían más a la tradición liberal del empirismo y la escuela escocesa de sentido común que a la más restrictiva tradicional en el pensamiento alemán. En Norteamérica, la posición de James (1890), en los *Principles*, había reflejado una actitud también más abierta con relación a las posibilidades de la introspección que la que tenía Wundt («A la observación introspectiva —decía James— debemos atenernos primera y principalmente y siempre») y Titchener no fue fiel, en realidad, al enfoque restrictivo del maestro alemán, sino más bien a las tradiciones liberales de la filosofía empirista.

La posición de Wundt en relación a los experimentos que dieron lugar a la célebre polémica del «pensamiento sin imágenes» es muy ilustrativa del enfrentamiento sobre los dos modos de entender el valor de los enunciados mentales en primera persona, a los que nos hemos referido en este apartado. En 1907, Wundt publicaba su crítica a los discí-

pulos de Külpe, en un artículo titulado «Acerca de los experimentos de *ausfrage* y de los métodos de la psicología del pensamiento», en que reivindicaba la necesidad de volver al enfoque restrictivo.

Los trabajos de Bühler, por ejemplo, que habían conducido a las posiciones más radicales de los teóricos del pensamiento sin imágenes, no eran verdaderos experimentos científicos por varias razones: (1) no permitían diferenciar sujeto y objeto de observación, (2) ni aseguraban las condiciones de repetibilidad de las observaciones, (3) ni permitían la percepción interna de lo mental, al plantear la necesidad de compatibilizar esa percepción con la resolución de problemas complejos. No eran los de Bühler verdaderos experimentos científicos, sino «autoobservaciones en condiciones difíciles que favorecen más el autoengaño que la autoobservación» (Wundt, 1907, p. 344). Así, no era extraño que esos pseudo-experimentos terminasen por caer en una especie de «dualismo», por encontrar formas, en el pensamiento, de «espíritus sin contenidos», parecidos a los «actos puros» de los escolásticos.

No revisaremos aquí la polémica del pensamiento sin imágenes (vid. Humphrey, 1973, Carretero y García Madruga, 1984), que se centró en un tema básico para los enfoques cognitivos: el de la naturaleza de las representaciones mentales en las funciones de pensamiento. Baste con recordar que aquella polémica jugó un papel decisivo en el descrédito de la psicología introspectiva (Watson aludió explícitamente a ella en el artículo inaugural del conductismo, de 1913). En los términos analíticos que hemos utilizado a lo largo de este capítulo, podríamos decir que el drama de aquella psicología fue, quizá, el de confundir un sistema destinado a la previsión de la conducta propia y ajena con un mecanismo de acceso a una maquinaria mental supuestamente transparente.

Esta «ilusión peligrosa de la psicología» era ya intuida y anticipada implícitamente por Wundt y James, cuando hablaban, por ejemplo, del error de objeto, de la tendencia a confundir los objetos que se presentan a la conciencia con la conciencia misma de tales objetos.

La duda que seguimos teniendo, y que no se ha despejado en el curso de la historia de nuestra ciencia, es la de si esa confusión no será acaso inevitable. Si no será que la conciencia sólo lo es de los objetos y de atributos muy limitados de nuestras relaciones con ellos, y no, en absoluto, de los procesos y representaciones (simbólicas o no) que nos permiten precisamente definir aquellos objetos y estas propiedades.

10.3. **La reducción extensional del conductismo**

Si el drama de la Psicología Introspectiva consistió, quizá, en tomar un instrumento de predicción psicosocial por un aparato de observación del mecanismo de la mente, el del conductismo fue el de no haber podido lograr nunca (con la única excepción quizá de la obra de Skinner) su gran ambición: la de hacer una psicología cuyos enunciados empíricos *y teóricos* estuvieran compuestos exclusivamente de términos *extensionales.* La peculiaridad de la posición conductista puede sintetizarse en pocas palabras: trataba de basar la psicología en enunciados acerca de la conducta estrictamente simétricos y en tercera persona de singular, es decir, acerca de «un otro», concebido como un objeto infinitamente distante de la subjetividad propia. Trataba de formular explicaciones que no contuvieran ninguna clase de entidades mentales, ninguna clase de entidades definidas por esas dos notas de interioridad e intencionalidad, que —como hemos visto en otro momento— son los rasgos esenciales por los que se define lo mental.

El problema es que, desde un principio, desde los propios escritos fundacionales de Watson (1913, 1914, 1919, 1924), la posición conductista partió de ciertos equívocos y de ciertos juegos de malabarismo conceptual que permitieron la inclusión de notas de intencionalidad *malgré lui*, de forma indirecta y no reconocida.

Esto lo ha visto muy bien Yela (1974), en relación con la noción de «estímulo». El estímulo es, sí, la energía física, y por ello puede definirse de forma estrictamente extensional, como se define cualquier otra energía. Pero el estímulo es, también, aquello a lo que responde el organismo. Y el organismo no responde, de forma directa, a las energías físicas, sino que responde a objetos significativos. «Es correcto —dice, por ejemplo, Watson— hablar de una masa total de factores estimulantes que, como un todo o una situación, producen la reacción del hombre ante ella» (1919, p. 10) y, pocas páginas más adelante, pone como ejemplo de estímulo «el hecho de que ustedes estén aquí, en una sala de conferencias».

Pero, para el organismo que responde a ello, «el hecho de que ustedes estén aquí» es un objeto intencional. Los objetos y las situaciones a los que responden los organismos son objetos intencionales. Los organismos no responden a objetos tales como las ondas o cuantos de luz, las vibraciones de las moléculas del aire y las sustancias químicas que son registradas por los receptores del olfato. O, mejor dicho, sólo res-

ponden a tales objetos de forma muy indirecta, después de procesos muy complejos de elaboración. De forma directa responden a lo que «significan» las situaciones para ellos. A estímulos que son —todo lo más— estímulos *distales.*

Como muy bien veía Brunswick (1937), el estímulo es una parte del ambiente ecológico que se hace objetivamente presente a un sujeto (*intentional erreicht).* Este responde, es verdad, a los cambios en las ondas (o los cuantos) de luz, y en las vibraciones del aire que producen el sonido. Pero su conducta no se explica como respuesta a esas transformaciones moleculares y extensionales, sino como respuesta a objetos *significativos*, que es como decir *intencionales.* Es decir, como respuesta a las complejas elaboraciones que realiza un sistema cognitivo de los estímulos proximales que son transducidos por sus receptores.

Esto mismo puede decirse del otro concepto fundamental de los conductistas: la *respuesta.* En una definición puramente extensional, la respuestas consisten en movimientos (que pueden describirse hasta el milímetro). Pero no son ésas, en rigor, las respuestas que interesan al psicólogo. A los profesionales de la mente nos interesan las respuestas como unidades de acción significativa, y las respuestas a las que no pueden asignarse significados nos importan más bien poco. Es quizá posible agotar la definición de la reacción de estornudar mediante una descripción minuciosa de los reactivos que la producen y de los cambios en el sistema respiratorio que dan lugar a ella. Pero las respuestas que está produciendo quien escribe estas páginas, por ejemplo, no quedan descritas psicológicamente cuando se define cómo bajan los dedos hasta percutir en las teclas del teclado que usa para escribir, ni al señalar las complejas adaptaciones articulatorias que se ponen en juego en la producción de la escritura como movimiento. La respuesta de escribir no consiste sólo en sus movimientos. Es, también, una conducta intencional. Su descripción tiene que incluir el hecho de que se escribe algo, y su carácter de' respuesta psicológica depende de lo que se escribe y de un entramado significativo de propósitos y de significados con los que se escribe.

El rechazo de Watson a la introspección y la conciencia se acompañó de un abandono de las nociones representacionales que habían servido, en la psicología anterior, para explicar las funciones superiores de pensamiento y lenguaje. Esas nociones (imágenes, formas internas, *Bewusstseinslagen*, etc) poseían las notas de interioridad e intencionalidad de lo mental, y no cabían en el intento de reducción extensional de Watson.

Si la primera psicología experimental fue, en palabras de Humphrey, una «psicología de la representación», parecería que la nueva psicología no-mentalista era sencillamente una «psicología sin representación».

Sin embargo, la operación teórica de Watson consistió, en realidad, en tratar de mantener conceptos de representación, pero despojados de su naturaleza intencional e inevitablemente interna (con arreglo a su noción operacionalista y materialista de la naturaleza de los conceptos científicos). Para ello, se sirvió de una versión esencialmente *eferencial* de las representaciones. La representación tendría un significado muy literal: la nueva presentación de una respuesta. Las representaciones no serían conocimientos, sino *respuestas.* La representación no es, para Watson, una re-instauración de una sensación o percepción originaria, sino de un «acto muscular»:

> Creo —decía Watson— que se puede mantener la hipótesis de que los llamados «procesos superiores de pensamiento» se desarrollan en términos de un tenue restablecimiento de un acto muscular originario (incluyendo a la palabra entre los actos musculares) y que se integran en sistemas que responden con arreglo a un orden serial (mecanismos asociativos). Con ello, los procesos reflexivos se convierten en algo tan mecánico como los hábitos (1919, p. 86).

En ese comentario se sintetizan aspectos esenciales de la primera posición conductista con relación a las funciones superiores. Esta posición se caracteriza por los siguientes aspectos:

(1) Reducir esas funciones a movimientos y estímulos, cuando menos potencialmente observables,

(2) relacionados de forma *serial,* en una cadena asociativa de izquierda a derecha, de una sola dimensión «plana», y

(3) que funcionaría de forma mecánica, tan mecánica «como los hábitos».

Los criterios de **observabilidad, mecanicismo** y **determinación causal y eficiente de lo posterior por lo anterior,** que presidieron la selección conductista de los conceptos psicológicamente útiles, se manifiestan en el comentario citado de Watson con toda claridad. En esencia, estos criterios siguieron presidiendo después la selección de constructos teóricos por los representantes del conductismo mediacional, que se basarían en la posición de Hull. Para éste:

> el mecanismo rg---sg (de respuestas fraccionales mediacionales y estímulos propio-ceptivos provocados por ellas) conduce, de una forma estrictamente lógica,

a lo que se consideraba en principio el núcleo mismo de lo psíquico: el interés, el planeamiento, la previsión, la presciencia, la expectancia, el propósito, etc» (1952, p. 350).

Las elaboraciones más complejas del conductismo mediacional de Osgood, Staats, Mowrer y otros, se sitúan en este mismo marco teórico. Así, por ejemplo, Osgood empleó las concepciones de Hull, sobre la respuesta fraccional anticipatoria de meta, para desarrollar su teoría mediacional del significado y sus cada vez más complejas elaboraciones posteriores (vid. Osgood, Sucy y Tannembaum, 1957). El significado (que es, en definitiva, otro modo de designar el concepto de «intencionalidad») se definía, en aquellos modelos sucesivos, en términos de cadenas cada vez más largas y tortuosas de estímulos y respuestas de carácter interno y mediacional, pero que seguían manteniendo la forma estrictamente extensional, y respondiendo a los principios de formación (por procesos de condicionamiento y asociación) que permitían justificar los conceptos teóricos de los modelos- conductistas. El intento tardío de Osgood y Hoosian (1974), de identificar las respuestas mediacionales con rasgos semánticos, trataba de alcanzar una solución de compromiso entre las nuevas posiciones cognitivas, que incluían nociones caracterizadas por notas de estructura formal, referencia intencional e interioridad, y los viejos conceptos conductistas que daban cuenta del significado (es decir, del núcleo de lo intencional) en términos de cadenas asociativas E-R de izquierda-derecha.

Desde la posición del conductismo radical, el intento de Skinner (más consecuente y, a mi entender, más coherente y fértil que los anteriores) implicaba una posición de no compromiso con ninguna clase de conceptos mentalistas, por mucho que se camuflasen. Esta actitud se manifestó, por ejemplo, en el intento de explicar los fenómenos de significado del lenguaje en términos, de nuevo estrictamente extensionales, de relaciones funcionales entre estímulos antecedentes, o consecuentes, y conductas operantes. Este enfoque ha tenido históricamente consecuencias positivas: ha permitido desarrollar, partiendo de los desarrollos del análisis experimental de la conducta, modelos útiles de intervención sobre el lenguaje; además, ha contribuido a explicar la naturaleza funcional de la actividad lingüística.

Sin embargo, como es bien sabido, del mismo modo que las críticas a los psicólogos introspectivos implicados en la polémica del pensamiento sin imágenes fueron determinantes en el origen del conductismo, las

realizadas por Chomsky al uso peculiar de conceptos extensionales (como el de «fuerza» de la respuesta) en *Verbal Behavior,* jugaron un papel emblemático en los comienzos de la Psicología Cognitiva. En su crítica de **Verbal Behavior,** Chomsky (1959) denuncia el uso puramente metafórico y completamente inapropiado de los términos «teóricos» del conductismo radical (como el de «fuerza de respuesta») en la explicación de la actividad humana más prototípicamente significativa: la actividad lingüística.

En la perspectiva de análisis que estamos utilizando, resulta interesante destacar el hecho de que uno de los artículos emblemáticos de la nueva psicología que se anunciaba en los años sesenta, la psicología cognitiva, consistiera precisamente en una demostración de la imposibilidad e inadecuación del empleo de un vocabulario puramente extensional para explicar una función mental que constituye el paradigma de la intencionalidad. La propuesta positiva, que subyacía a la crítica de Chomsky a Skinner, era la de volver al empleo de términos intencionales, pero revistiéndolos de la rigurosa definición mecanicista de las teorías de autómatas. Objetos tales como las estructuras profundas, entendidas como formas intencionales de representación mental, y que se derivarían, de forma precisa y automática, mediante la aplicación de reglas lingüísticas bien definidas.

Pero ésa ya es otra historia, a la que debemos referirnos brevemente en el punto siguiente: la historia del difícil intento cognitivo de integrar, en unos mismos patrones explicativos, el mecanicismo abstracto que había permitido el nacimiento de las nuevas tecnologías del conocimiento y la vieja intuición de la mente como sistema intencional.

10.4. El origen del paradigma C-R: una mecánica de la intencionalidad

En algunos trabajos históricos, se han analizado, con cierta profundidad, las condiciones en que surgió la psicología cognitiva y los factores que contribuyeron a su primer desarrollo (vid., por ejemplo, Gardner, 1985). No haremos aquí un análisis histórico pormenorizado de esas condiciones y factores, sino que emplearemos los conceptos analíticos que hemos ido desarrollando a lo largo de los capítulos anteriores para definir brevemente algunas características de las explicaciones cognitivas que parecen importantes, y que no habían sido vistas con anterioridad.

Una marca de las que definen los cambios de paradigmas en las ciencias sociales y de la conducta es el hecho de que, en esos cambios, se modifican *los criterios de definición de los conceptos admisibles como objetos teóricos.* Así, los objetos teóricos del conductismo se justificaban de diferente modo a como se justificaban las entidades explicativas de las psicologías introspectivas, en las que la evidencia consciente de los fenómenos de la mente constituía el criterio último de justificación.

Frente al criterio de justificación introspectivo, los conductistas avalaban sus constructos teóricos en criterios tales como:

(a) La continuidad en los mecanismos de formación de las entidades internas (por ejemplo, los estímulos y respuestas mediacionales) y las externas; esos mecanismos de formación serían, en general, procesos asociativos y de condicionamiento operante o clásico.

(b) La observabilidad potencial de los referentes de tales constructos (recordemos los intentos de Watson de medir, con un «galvanómetro de cuerda» las respuestas de hioides, labios y lengua, mientras sus sujetos pensaban).

(c) La observabilidad indirecta (por ejemplo, en los trabajos de Sokolov, 1972, y MacGuigan 1978, sobre las manifestaciones sub-conductuales del pensamiento).

Pues bien, un aspecto esencial que define el origen de la Psicología Cognitiva consiste, de nuevo, en un cambio en los criterios de justificación de las entidades teóricas: ahora estas entidades se justifican en tanto en cuanto se integran en un sistema de cómputo (en un algoritmo) que simula la conducta que se quiere explicar. Los esquemas, las proposiciones, los conceptos se admiten porque se engranan en una u otra «maquinaria abstracta», de naturaleza computacional y permiten explicar objetivamente observaciones extensionales (y mensurables) sobre la conducta de sujetos situados, generalmente, en condiciones experimentales muy precisas.

Los datos de observación no son, por consiguiente, más subjetivos en la psicología cognitiva de lo que puedan serlo en el conductismo. Son completamente objetivos: el tiempo que tardan las personas en responder a problemas, en decidir si una secuencia de letras es una palabra o no, etc; también los errores que cometen cuando razonan, o cuando producen lenguaje, etc. Son datos que pueden situarse perfectamente en escalas métricas y, generalmente, someterse a un tratamiento estadístico riguroso. Son, en definitiva y en una palabra, datos estricta y escuetamente extensionales.

En este punto, los análisis históricos han caído frecuentemente en ciertos equívocos. La interpretación tradicional, según la cual las restricciones teóricas impuestas por el conductismo llegaron a hacerse insostenibles, en los años cincuenta, y tuvieron que ser transgredidas —para dar cuenta de hechos tales como la influencia de las imágenes en el aprendizaje verbal, etc—, es una interpretación que tiene algo de cierto, pero que no agota toda la verdad de la historia.

La Psicología Cognitiva implicó, sí, una recuperación del vocabulario mentalista que había sido proscrito previamente desde la filosofía conductista. En ese sentido representó una cierta liberación, una liberalización aparente de los criterios de admisión de los constructos teóricos. Después de varios años de dieta conceptual, parece que los psicólogos podían hablar de nuevo de imágenes y planes, significados y pensamientos...de estructuras profundas incluso. Sin embargo, ello no significó una simple disolución de los criterios justificativos anteriores, sino, sobre todo, su *sustitución parcial* por otros criterios no menos estrictos.

¿Por qué hablamos de sustitución «parcial»? Para explicarlo, debemos señalar que, como han indicado algunos historiadores de la psicología (por ejemplo, Leahey, 1981) los rasgos de continuidad entre la psicología cognitiva y el conductismo son más significativos de lo que se ha supuesto tradicionalmente. Leahey realizó, en 1981, una comunicación a la A.P.A con un título bastante provocativo: «La revolución que nunca existió: el procesamiento de la información *es* conductismo». No está el autor de acuerdo con el contenido de ese título, pero sirve para destacar la necesidad de atender a aquellos aspectos, muy importantes, en que la psicología cognitiva supone, en efecto, una continuación de enfoques metodológicos y de contenidos que habían mantenido previamente los conductistas.

¿Qué aspectos son ésos? A mi parecer, dos muy importantes:

(1) La psicología conductista se basó, desde los escritos inaugurales de Watson (1913) en la premisa de que los enunciados observacionales, de que debe partir la ciencia psicológica, tienen que ser extensionales y en tercera persona de singular o plural (de ningún modo, en primera). No viene al caso ahora si tales enunciados se refieren a la «fuerza» o intensidad de una respuesta, a su frecuencia, a su latencia, a que sea categorizada como acertada o errónea con relación a un criterio externo. Lo que importa es que los *datos* con que cuenta el psicólogo son de carácter extensional. Y este principio se ha mantenido, en lo esencial, en Psicología Cognitiva.

En algunas áreas de estudio (como la solución de problemas) se han utilizado, desde ciertas líneas definidas de investigación, informes de naturaleza introspectiva (Newell y Simon, 1972; Ericsson y Simon, 1984). Pero estos casos son la excepción más que la regla, y suelen compensarse, además, con una acentuación de las exigencias sintácticas de los modelos. Además, curiosamente son más frecuentes entre investigadores que tienen su origen en la Inteligencia Artificial, y muy raros en los formados como psicólogos. En general, los psicólogos cognitivos hemos heredado las suspicacias conductistas con respecto al valor de los enunciados mentales en primera persona.

Los enunciados observacionales que configuran la «base semántica» de las teorías cognitivas son, como los conductistas, extensionales. La diferencia entre los enunciados conductistas y los cognitivos es más de matiz: aquéllos empleaban, preferentemente, dimensiones como la intensidad o la frecuencia para la definición operacional de las conductas estudiadas; éstos utilizan, sobre todo, los tiempos de respuesta (Lachman, Lachman y Butterfield, 1979).

(2) El segundo aspecto en que sé produce una continuidad (muy matizable) entre la psicología cognitiva del procesamiento de la información y el conductismo es en la nota de *mecanicismo* que caracteriza a ambas posiciones. Sin embargo, en este aspecto, las diferencias son mucho más fundamentales. El modelo de máquina que subyacía a las posiciones conductistas correspondía a un enfoque E-R, basado en un paradigma mecánico que no es el que utiliza la Psicología Cognitiva. El macanicismo conductista era heredero de las interpretaciones de la mecánica clásica, en que la «información» no jugaría ningún papel explicativo. Por el contrario, como hemos comentado más extensamente en el capítulo segundo, los enfoques C-R y de Procesamiento de la Información, clásicos de la Psicología Cognitiva, están comprometidos en realidad con una posición de mecanicismo abstracto, que remite a autómatas simbólicos que tienen poco que ver con el tipo de mecanismos que formaron parte de las imágenes de los primeros conductistas.

De las dos observaciones anteriores se derivan los que son, a mi entender, los criterios de justificación básicos de la psicología cognitiva, y a los que ya hemos hecho referencia a lo largo de estas páginas:

(1) La base observacional de las teorías cognitivas se compone de observaciones sobre aspectos extensionales de la conducta de las personas: cuánto tardan en resolver el problema X, o cuántos errores cometen en él, qué enunciados recuerdan y cuáles no, cuánto tardan en comprender un enunciado, etc. Lo que las personas dicen sobre sus funciones mentales puede tener, en algunos trabajos, un papel complementario, pero no es, casi nunca, el fundamento de las construcciones teóricas.

LLamaremos a éste *criterio de extensionalidad de la base empírica.*

(2) El segundo criterio es el que se refiere al mecanicismo. Viene a ser el siguiente: los objetos teóricos de los psicólogos cognitivos se justifican en tanto en cuanto se insertan en *algoritmos,* o *procedimientos efectivos,* capaces de generar las mismas conductas que se definen extensionalmente en la base teórica (y, con un criterio de adecuación explicativa fuerte, de generarlas con procedimientos esencialmente idénticos a los que emplean las personas, cuando se enfrentan a las mismas tareas que los algoritmos).

Llamaremos, a este segundo, *criterio de justificación algorítmica de los conceptos teóricos.*

La ruptura «mentalista» de la Psicología Cognitiva del procesamiento de la información sólo puede entenderse cuando se sitúa en relación con esos dos criterios. La recuperación de los constructos explicativos que implican conocimiento (que se habían perdido en la travesía conductista) sólo fue posible en función de esas dos prescripciones:

(1) Explicar la conducta, y

(2) explicarla en términos mecánicos.

Sólo así fue posible la tercera prescripción:

(3) explicarla como producto del conocimiento.

La recuperación de los viejos conceptos de la psicología natural mentalista, de los planes y propósitos (Miller, Galanter y Pribram, 1960), las estrategias activas e internas de elaboración del conocimiento (Bruner, Goodnow y Austin (1956), los símbolos mentales, como objetos intencionales y portadores de conocimiento (Newell, Shaw y Simon, 1958) estuvo avalada por la doble justificación ofrecida por la naturaleza extensional de los datos y el carácter supuestamente mecanicista de las explicaciones.

Estos criterios de justificación no siempre se han cumplido. Más aún: el segundo se alcanza raras veces y, como ya hemos comentado en otro punto, las explicaciones cognitivas no suelen llegar a alcanzar el grado de precisión que se obtiene cuando se construyen algoritmos completos, capaces de generar todo el campo de conductas estudiado. La gran mayoría de los experimentos cognitivos lo que hacen es demostrar la influencia de ciertos *factores* en los procesos de razonamiento, solución de problemas, comprensión del lenguaje, memoria, etc, pero no permiten

construir procedimientos efectivos completos capaces de simular explicativamente las conductas.

En ese sentido, es preciso distinguir entre los criterios que se aplican a las investigaciones cognitivas concretas y aquellos por los que se justifica la psicología del procesamiento de la información como un todo: El criterio algorítmico suele ser, de hecho, un *ideal,* de forma parecida a como fue un ideal del conductismo la reducción de los procesos de desarrollo del pensamiento y el lenguaje a los mecanismos asociativos relativamente simples encontrados en el aprendizaje y el condicionamiento animal.

El supuesto algorítmico, que establece que las funciones mentales (mientras no se demuestre lo contrario) pueden reducirse a procedimientos efectivos, que consisten en la aplicación de procesos elementales a representaciones moleculares, es un postulado terminal que justifica la empresa global de la psicología cognitiva, pero que no suele concretarse en las explicaciones concretas de conductas específicas.

Estas consideraciones sirven para explicar la significación de las primeras críticas cognitivas a las formas de explicación de los conductistas, y las primeras propuestas de una psicología de estilo C-R, o de Procesamiento de la Información. Por ejemplo, en el importante simposio de Hixon en 1948, Karl Lashley explicó que los modelos explicativos conductistas «de izquierda a derecha» (que eran una consecuencia inevitable de la epistemología ingenuamente «realista» , que sólo podía concebir lo real como lo que se expresa materialmente en una secuencia temporal) eran incapaces de dar cuenta de conductas organizadas, reclamando la necesidad de recurrir a organizaciones o estructuras jerárquicas para explicar esas conductas.

En aquella ocasión, Lashley, que había estado muy cerca del conductismo, fue explícito en la reivindicación de la importancia del lenguaje como foco de estudio de los sistemas de organización jerárquica que dan lugar a la conducta: «Los problemas que crea la organización del lenguaje —decía— me parecen típicos de casi cualquier otra actividad cerebral» (cit. por Gardner, ed. esp. 1987, p. 28). Era éste uno de los primeros intentos de convertir al lenguaje en paradigma de explicación, en vez de explicar el lenguaje, como anteriormente, por reducción a funciones más elementales. Como señala Gardner:

Lashley adoptó una posición radical, tanto por los temas que abordó como por la forma en que lo hizo. Los científicos que se ocupaban del comportamiento humano se habían mostrado renuentes a investigar el lenguaje, a causa de su complejidad y de su relativa «invisibilidad» como forma de conducta; y cuando lo hicieron, lo típico fue que buscaran analogías entre él y otras conductas más simples (como correr por un laberinto o picotear trozos de comida en una jaula) de organismos también más simples (ratas o palomas). Lashley no sólo se centró en el lenguaje, sino que mostró su deleite por la complejidad de éste, e insistió en que otras actividades motrices no eran menos intrincadas (1987, pp. 29-30).

La posición de Lashley, en la reunión de Hixon, prefiguraba aspectos importantes de lo que luego había de ser la actitud cognitiva que se desarrollaría en los años siguientes: la consideración del lenguaje como foco de interés y de los modelos lingüísticos como modelos explicativos ejemplares, la insistencia en mecanismos de determinación de arriba a abajo, y de carácter jerárquico, el gusto por la complejidad, iban a definir algunas de las características de aquella actitud, que se puso claramente de manifiesto ocho años después, en las jornadas sobre Teoría de la Información, celebradas en 1956 en el M.I.T.

En aquella reunión, se presentaron algunas de las ideas que iban a tener más influencia en la germinación de los enfoques cognitivos. Noam Chomsky concretaba en el estudio del lenguaje, de forma rigurosa, algunas de las intuiciones adelantadas por Lashley, al demostrar que las gramáticas de estados finitos empleadas por los teóricos de la información eran incapaces de generar el lenguaje. El modelo «generativo-transformacional», que él proponía, se convertiría, posteriormente, en el ejemr piar prototípico de los enfoques explicativos de determinación jerárquica. La idea de que las explicaciones psicológicas deben consistir en teorías de cómputo sobre representaciones simbólicas y en modelos algorítmicos capaces de generar el campo potencial de conductas de interés se prefiguraba también en la presentación de Chomsky.

Aquella idea era un resultado de las mismas reflexiones sobre lenguajes formales que se encarnaban, por aquellos años, en los primeros ordenadores (para este aspecto de la historia, vid. Bretón, 1989). En la misma reunión en que Chomsky presentó su aportación en que demostraba la imposibilidad de generar el lenguaje con una máquina que actuara con arreglo a un proceso de determinación probabilística de izquierda a derecha, Newell y Simón describían la «máquina de la teoría lógica», que constituía la primera demostración completa de un teorema llevado a cabo por una computadora. Y, en aquella reunión, George Miller ponía de manifiesto, de forma muy brillante, algunas limitaciones

de los sistemas de procesamiento, en su presentación sobre el mágico número 7 más-menos 2.

Este recuerdo rápido de la reunión «inaugural» de la psicología cognitiva en el M.Í.T. no tiene aquí ninguna pretensión histórica (vid. Gardner, 1987, para ese aspecto) sino simplemente la de traer a nuestra reflexión las ideas germinales que definieron, en una primera fase, el desarrollo de los modelos cognitivos. En el mismo año 1956, Bruner, Goodnow y Austin desvelaban, en *A Study of Thinking,* las estrategias activas de elaboración de la información que los sujetos emplean en el laboratorio en procesos de formación de conceptos clásicos. Y un año después, en 1957, Chomsky presentaba, en *Syntactic Structures,* el primer modelo generativo transformacional de la competencia lingüística. La influencia de ese modelo lingüístico de explicación se reconocía explícitamente en *Plans and the Structure of Behavior,* de Miller, Galanter y Pribram (1960), y la formulación del modelo simbólico-computacional de explicación se hacía claramente explícita en el importante artículo de Newell, Shaw y Simon (1964), «Elements of a Theory of Human Problem Solving».

Conviene que nos detengamos algo en una formulación integrada de lo que significaban las propuestas de aquellos trabajos, porque constituyen el marco de referencia del desarrollo posterior de los modelos cognitivos.

El intento explicativo esencial que se formulaba en todas estas primeras aportaciones a la psicología cognitiva naciente era el de definir sistemas de cómputo de gran poder. Se suponía que esos sistemas de cómputo permitirían explicar la generación de estructuras bien definidas (por ejemplo, oraciones gramaticales) o el enfrentamiento a tareas bien delimitadas de formación de conceptos y solución de problemas, por medio de estrategias activas de elaboración del conocimiento.

Las notas de actividad, estructura, determinación jerárquica, dependencia formal de los procesos, y naturaleza simbólico-computacional de éstos, que han definido esencialmente a los modelos cognitivos, se encontraban ya prefiguradas en aquellos escritos inaugurales de la Psicología Cognitiva. El supuesto subyacente fundamental era el de que es posible dar cuenta de las funciones humanas complejas, tales como las de pensamiento y lenguaje, por medio de la reconstrucción de la naturaleza de una maquinaria mental que transforma símbolos y trata con conocimientos. Es evidente que el modelo de la Máquina de Turing, al que nos referíamos más extensamente en el capítulo segundo, ocupaba una

posición central (aunque no siempre explícita) en aquellas primeras formulaciones de la psicología cognitiva.

Hay dos aspectos de esas primeras formulaciones del nuevo plano de explicación cognitiva, que merecen algún comentario:

(1) Desde un principio, las nuevas formulaciones se entretuvieron en un cierto equívoco que, a mi entender, ha sido enormemente influyente en el desarrollo y la «imagen social» de nuestra ciencia.

Este equívoco es señalado muy acertadamente por Kintsch, Miller y Polson (1984), y consiste en lo siguiente: se mezclan los conceptos propositivos e intencionalistas de la «psicología natural», a los que nos hemos referido a lo largo de estas páginas, con las categorías derivadas del modelo C-R de explicación. Nociones que sólo tienen sentido en un nivel *personal* de descripción, como las de «plan» y «propósito», o que están inevitablemente ligadas a la fenomenología de la conciencia, como la de «imagen», se entreveran, en las mismas explicaciones, con nociones como la de cómputo o sistema TOTE, que sólo pueden tener sentido en un nivel sub-personal de explicación.

Esta tendencia se expresa con especial claridad, a mi entender, en *Plans and the Structure of Behavior,* de Miller, Galanter y Pribram (1960). Quizá por razones de claridad expositiva (y Miller es, desde luego, un maestro en eso), los autores recurren constantemente a ilustraciones, muy expresivas, de sus ideas. Estas ilustraciones se vinculan a las intuiciones intencionalistas que tenemos como psicólogos naturales. Nos parece que hablan de los «planes» que uno tiene en la cabeza, y de los que es consciente, de las imágenes que nos vienen a la mente, etc.

Sin embargo, la aportación fundamental que hacen Miller et al. (1960) consiste en definir la forma general (i.e. la organización TOTE) de los mecanismos subpersonales relevantes para estudiar los procesos de conocimiento.

(2) Las primeras explicaciones situadas ya claramente en el nuevo plano cognitivo conservan la huella del viejo intento de encontrar un «lenguaje universal del pensamiento» o, (con una expresión redundante pero útil), un «lenguaje universal del lenguaje», que sería un sistema único de cómputo, el cual, partiendo de un conjunto limitado de principios , y de elementos simbólicos y procesuales, daría cuenta de ámbitos completos de la actividad humana: la producción de oraciones gramaticales, en el caso de la gramática generativo-transformacional, la solución humana de problemas, en el del G.P.S (El «Solucionador General de Problemas») de Newell y Simon, etc.

Es decir, los primeros modelos explicativos, de los que nació la Psicología Cognitiva, se basaron generalmente en el supuesto de que *es posible definir mecanismos muy generales de inferencia, generación lingüística o solución de problemas, cuyo funcionamiento estaría determinado por la forma.* Este supuesto estaba ligado al metapostulado logicista (De Vega, 1981), que formaba parte de sus presuposiciones teóricas. En general, las primeras explicaciones cognitivas fueron «explicaciones sintácticas», y llevaron a la realización de experimentos que trataban de desvelar, sobre todo, el papel de los factores sintácticos en las funciones de Pensamiento y Lenguaje.

En el estudio del lenguaje, por ejemplo, las investigaciones pioneras de la psicolingüística de enfoque cognitivo se centraron en el intento de dar identidad psicológica a las reglas formales de transformación (del todo ajenas, en el primer modelo de la gramática generativo-transformacional, al componente semántico). Este era el intento que perseguían Miller y McKean (1964) en su investigación sobre la transformación pasiva, Savin y Perchonock (1965), en su investigación sobre el recuerdo de oraciones pasivas, negativas e interrogativas, Mehler (1963) en su comprobación de la tendencia de los sujetos a «simplificar» transformaciones, eliminándolas en el recuerdo, etc.

Sin embargo, ya desde mediados de los años sesenta, empezaron a reclamar su papel explicativo los factores semánticos y de contenido, que llevaron finalmente al abandono de la hipótesis de que las dificultades de comprensión constituyen un espejo sin mancha de la complejidad derivacional (entendida en términos sintácticos) de las oraciones. (Wason, 1965; Greene, 1970; Johnson-Laird, 1968; Clark, 1965; Johnson, 1967).

Algo semejante sucedió en los estudios de pensamiento. Los modelos sintácticos más ambiciosos y generalistas, como el «Solucionador General de Problemas» de Newell y Simon (1972), demostraron pronto que su ámbito explicativo era mucho más restringido de lo que se pretendió en principio, y que sólo eran capaces de explicar la solución de problemas cerrados y de poca «densidad semántica» (Wason, 1978). Aunque el GPS fuera capaz de simular algunas conductas humanas de solución de problemas, su ejecución en ciertos casos era muy cercana a la de ensayo y error, y se mostraba incapaz de improvisar estrategias defectivas o heurísticos en su aplicación. Como ha señalado Asensio (1990):

El supuesto básico ele la metáfora del ordenador orientó a los modelos del procesamiento de la información hacia la consideración de la mente humana como un procesador de cálculos sintácticos, donde el soporte lógico venía definido por el cálculo booleano. Este planteamiento resultó extremadamente útil para analizar las restricciones estructurales de los procesos psicológicos. Sin embargo, uno de los problemas fundamentales con los que se encontraron, desde el primer momento, estos estudios consistió precisamente en cómo dar cuenta, desde esta perspectiva exclusivamente lógico formal, de los contenidos de la mente, ya que existen numerosas evidencias experimentales (que veremos en el próximo capítulo) de la influencia del contenido temático en las tareas cognitivas (p. 149).

En los estudios de razonamiento, las influencias de los contenidos en los errores cometidos en ciertas clases de silogismos (en especial, en los silogismos categoríales) se conocían desde varios años antes de la «promulgación» explícita del paradigma cognitivo (Lefford, 1946; Janis y Frick, 1943; Morgan y Morton, 1944), y pronto se pusieron de manifiesto las limitaciones de los modelos «sintácticos» (como lo es, en cierto modo, el de Piaget), que proponen que los sujetos son capaces de emplear dispositivos formales y poderosos de inferencia (como el reticulado de las operaciones binarias interproposicionales, o el grupo INRC de Piaget), con independencia de los contenidos (Wason, 1966, 1968, 1969). Numerosos experimentos sobre la tarea de selección de Wason, por ejemplo, han demostrado que la presentación de esta tarea con contenidos «realistas» modifica, por completo, los porcentajes de soluciones correctas (Wason y Shapiro, 1971; Johnson-Laird, Legrenzi y Sonino-Legrenzi, 1972; Bracewell y Hidi, 1974; Van Duyne, 1974, 1976; Evans y Wason, 1976, D'Andrade, 1980; Griggs, 1982, etc).

En suma: el supuesto fundamental según el cual las funciones superiores P-L podrían explicarse mediante modelos sintácticos y universales, indiferentes a los contenidos, fue demostrándose como insostenible a lo largo de los años sesenta. Podemos decir, en un resumen muy rápido, que la Psicología Cognitiva de la década de los setenta se caracterizó por los siguientes rasgos:

(1) La complementación o sustitución de los modelos explicativos «sintácticos», dominantes aún en la década anterior, por modelos semánticos. Esta tendencia tuvo su expresión en psicolingüística a través de los estudios de comprensión —y, en grado mucho menor, de producción— centrados precisamente en la influencia de los factores semánticos y conceptuales. En el estudio del razonamiento y la solución de problemas, muchas investigaciones comenzaron a centrarse en el papel de los con-

tenidos sobre los procesos de pensamiento, y comenzaron a formularse *modelos semánticos* del pensamiento.

(2) El desarrollo de la propia psicología cognitiva hizo cada vez más necesario el recurso a nociones *macroestructurales* (y no sólo microestructurales) para explicar las funciones cognitivas. Esta tendencia, proveniente de la influencia convergente de las investigaciones y enfoques constructivistas sobre la comprensión y el recuerdo del discurso y de los intentos de desarrollar programas de Inteligencia Artificial capaces de comprensión, se expresa con la mayor claridad en la noción de *esquema.* Los fracasos en los intentos de formular programas de I.A., basados en un simple depósito léxico y unas reglas de inferencia o generación sintáctica y que fueran capaces de comprender o traducir (Winograd, 1981; Dehn y Schank, 1982) llevaron a la necesidad~de recurrir a macro-estructuras para explicar el hecho de que las funciones superiores a imitar no se guían solamente por las reglas formales de su sintaxis, sino por un amplio acervo de conocimientos acerca del mundo, organizados conceptualmente de forma compleja y delicada. Las nociones de «marco» (Minsky, 1975), «guión» (Schank y Abelson, 1977) y «esquema» (Rumelhart y Ortony, 1977) son expresiones de enfoques explicativos nuevos. Tratan de capturar propiedades del sistema cognitivo tales como la capacidad de integrar conocimientos en totalidades significativas, de anticiparse a informaciones nuevas, de añadir contenidos no establecidos de forma explícita en el texto o el discurso, etc.

(3) La tercera característica de los estudios cognitivos, desde la segunda mitad de los setenta, está muy relacionada con la anterior: consiste en la propensión creciente a realizar investigaciones sobre unidades estimulares cada vez más molares, y a reclamar la necesidad de «validez ecológica» de las investigaciones. Así, en el estudio del lenguaje, se mantienen los trabajos sobre comprensión oracional y léxica, pero crecen muy rápidamente los relacionados con la comprensión de unidades mayores de discurso y de texto (Kintsch, 1974). En solución de problemas y razonamiento, se emplean cada vez más, situaciones cotidianas o estudios sobre «pensamiento profesional», etc. (vid. Elstein y Bordage, 1984; Kuhn, Pennington y Leadbeater, 1984).

Podemos definir esta segunda etapa del desarrollo de la Psicología Cognitiva (que se extiende, aproximadamente, desde 1970 a 1985) como *fase semántica* (tal como lo hace Asensio, 1990). Entre sus productos más emblemáticos y significativos están, además de las teorías de esquemas, enfoques como el de los *modelos mentales,* de Johnson-Laird (1983),

que trata de explicar funciones de razonamiento deductivo y comprensión lingüística por procesos de base semántica (y, al mismo tiempo, estrictamente formalizables), o el *modelo de dependencia conceptual* de Schanck y Abelson (1977) que, renunciando explícitamente al intento de construir modelos I.A de la comprensión guiados por la sintaxis, parte de la premisa de que los procesos de comprensión lingüística están dirigidos por «guiones» y determinados por una organización conceptual, que define primitivos semánticos para los elementos léxicos. Una característica de estos modelos es que abandonan la pretensión anterior de simular las funciones humanas complejas mediante sistemas de inferencia independientes de los contenidos y de extremada generalidad.

Este proceso, de complementación y sustitución de los modelos sintácticos por los semánticos, ha tenido consecuencias importantes en la definición de los conceptos acerca de las funciones de pensamiento y lenguaje que tenemos los psicólogos cognitivos:

(1) Por una parte, nos ha llevado a cuestionar y limitar el papel de los modelos «canónicos» formales, al que me he referido en otro punto, en la explicación de los procesos mentales. Así, las teorías de esquemas (Wason, 1983; Rumelhart, 1980) y las llamadas «teorías temáticas» (Pollard, 1983) de explicación del razonamiento deductivo, proponen modelos en que el razonamiento no se guía por un sistema general de reglas de inferencia, sino por mecanismos dependientes de los contenidos. Y el modelo de la comprensión de Schank reduce el papel de la sintaxis en los procesos de comprensión a algunas actividades de complementación de los procesos de naturaleza semántica. Ello no quiere decir que en estos modelos desaparezca la intuición de «determinación por la forma». Lo que sucede es que las formas se amplían y se entrelazan con los contenidos, se diversifican y, en muchos casos, pierden la nítida y normativa claridad de las formas lógicas y matemáticas, que son exclusivamente sintácticas.

(2) Por otra parte, desde el momento en que estas investigaciones y modelos de los años 70 y 80 han puesto de manifiesto las estrechas relaciones entre formas y contenidos en los procesos cognitivos, han constituido un determinante fundamental de la necesidad de desarrollar *modelos de interacción en paralelo* de tales procesos. Los procesos sobre la forma no son independientes de los que se realizan sobre los contenidos (como demuestran, por ejemplo, los abundantes datos existentes sobre las ricas interrelaciones entre factores sintácticos y semánticos en los procesos psicolingüísticos). Ya en 1977, en el manual de Rumelhart so-

bre *Human Information Processing,* se avistaba esa exigencia que ha terminado por concretarse, a lo largo de la década posterior, en los modelos conexionistas del conocimiento (Rumelhart, McClelland y Grupo de Investigación PDP, 1986).

Las consideraciones anteriores pueden resumirse de forma muy concisa, empleando el tipo de lenguaje que venimos usando a lo largo de estas páginas: en el intento difícil de compromiso entre forma e intencionalidad, que siempre ha definido a la psicología cognitiva, predominaron en una primera fase las exigencias de la forma. Sin embargo, en todos los ámbitos de nuestra ciencia, el significado fue imponiendo progresivamente sus derechos: las influencias semánticas y pragmáticas en el pensamiento y el lenguaje demostraron, en seguida, que no podían ser ignoradas. La intencionalidad imponía una y otra vez su enigmática presencia en los fenómenos estudiados por los psicólogos cognitivos. Esa presencia se expresaba en formas tales como los efectos de los contenidos y de las «intenciones» en los procesos superiores de pensamiento y lenguaje, o la penetrabilidad de esos y otros procesos a factores de interpretación.

De este modo, la historia relativamente corta de la psicología cognitiva puede entenderse, de forma metafórica pero provechosa, como la historia de unas complejas relaciones entre intencionalidad y forma. O, si se quiere, como un indeciso enfrentamiento entre la ineludible exigencia de definición y claridad sintáctica que requiere cualquier explicación para ser considerada como propiamente computacional, y la inevitable necesidad de referencia intencional que requiere cualquier estado o proceso para ser considerado como propiamente psicológico. Sí, la historia de la psicología cognitiva clásica es, en buena medida, la historia de esta agonía de intención y forma. ¿O quizá no sean las relaciones intencionales otra cosa sino formas aún no bien comprendidas?; ¿cúal es la sintaxis de las relaciones entre los procesos mentales y sus contenidos? La historia de la psicología cognitiva tendrá que hacerse mucho más larga y esforzada antes de que podamos dar una respuesta, ni siquiera provisional, a estas preguntas inquietantes.

10.5. El conexionismo y la recuperación de los objetos extensionales como conceptos teóricos en psicología

¿Cómo pueden emerger los fenomenos intencionales, y las formas, de la mente a partir de un sistema estrictamente extensional, como lo

es el cerebro?; ¿cómo es posible que el cerebro «construya la mente» a través de su relación con el medio externo e interno? Estas preguntas han inquietado siempre a los psicólogos, y sus respuestas han sucumbido frecuentemente a la tentación del dualismo. Ya nos hemos referido a algunas formas diferentes que ha tomado esa tentación a lo largo de la historia. Para señalar que la mente (el pensamiento) es una sustancia simple, que no es conmensurable con las sustancias compuestas del mundo extenso, Leibniz acudía a la metáfora absurda de entrar en ella como en un molino, en el que sólo hallaríamos piezas que se impelen unas a otras, «pero nunca nada con lo que explicar una percepción». Por su parte, los teóricos del enfoque C-R recurrieron al expediente del dualismo funcionalista (basado en la imagen de la independencia entre programas y ordenadores) para justificar la posibilidad de estudiar los procesos y representaciones mentales con independencia de los fenómenos neurobiológicos. En realidad, el de algunos racionalistas era un dualismo ontológico y el de los modernos psicólogos cognitivos un dualismo metodológico, pero el efecto es, en uno y otro caso, la suposición de que es posible adentrarse en lo mental sin considerar el funcionamiento cerebral.

Sí, son muchos los que han terminado por pensar, sencillamente, que entre las intenciones y formas de la mente y la naturaleza extensional y material del cerebro habría una muralla lógica insalvable. Podemos encontrar numerosas citas que reflejan este modo de pensar, pero pueden bastarnos, por su contundencia y claridad, estas dos de Tindall, recogidas en los' *Principles of Psychology* de William James:

> Es impensable el paso de la física del cerebro a los correspondientes hechos de conciencia. Aun concediendo que un pensamiento definido y una acción molecular definida ocurran simultáneamente en el cerebro, no poseemos el órgano intelectual, ni al parecer ningún rudimento de tal órgano, que nos pudiera permitir pasar, mediante un proceso de razonamiento, de uno a otro (vid. James, ed. esp. de 1989, p. 120).

Y, en este otro párrafo:

> Podemos remontar el desarrollo de un sistema nervioso y correlacionar con él los fenómenos paralelos de sensación y pensamiento. Vemos, con certeza indudable, que van de la mano. Aunque nos lanzamos a un vacío en el momento mismo en que tratamos de comprender la conexión que hay entre ellos... No hay fusión posible entre las dos clases de hechos, ni energía motora en ei intelecto del hombre que lleve de uno a otro sin producir una ruptura lógica (ibídem).

El problema es cómo salvar esa ruptura lógica de que habla Tindall, entre el funcionamiento, por una parte, de miles de millones de pequeñas unidades neuronales que, sin duda, no tienen mente, ni producen representaciones definibles formalmente, ni poseen intencionalidad, y, por otra, la producción mental de pensamientos y otros objetos intencionales.

Esta vieja cuestión ha tenido, desde el nacimiento de la psicología cognitiva, dos respuestas diferentes:

(1) La más antigua y clásica es la del dualismo funcionalista: establece que la mente es semejante a una máquina universal y produce conocimiento (es decir, computa) con independencia funcional del soporte que permite que «corran» sus programas. Ya hemos comentado esta propuesta en diversos momentos de nuestra reflexión sobre los «objetos con mente».

(2) La respuesta más reciente es muy distinta: cuando dejamos de lado las propiedades no relevantes de la estructura y el funcionamiento cerebral (tales como el tamaño de las glías o el hecho de que la transmisión química del impulso nervioso se produzca gracias a la serotonina o a las catecolaminas) y nos fijamos sólo en un modelo abstracto del sistema nervioso como un **sistema de cómputo,** encontramos que emergen propiedades «mentales» relevantes del funcionamiento masivo, paralelo, y cooperativo-competitivo, de esas unidades y conexiones que en sí mismas no son inteligentes ni intencionales, y que se definen por características extensionales muy sencillas a las que ya hemos aludido en el capítulo cuarto: niveles de activación de las unidades y fuerzas de conexión entre ellas, umbrales de in-put y de out-put de dichas unidades, patrones de conexión, y ciertas reglas de propagación, activación y aprendizaje (Rumelhart, Hinton y McClelland, 1986).

Estas propiedades simples son suficientes, según los teóricos conexionistas, para dar cuenta de la capacidad que tendrían las redes neurales de producir objetos intencionales enormemente complejos, tales como los esquemas y conceptos (Rumelhart, Smolensky, McClelland y Hinton, 1986), y modos de funcionamiento muy semejantes a las de los procesos que se guían por reglas formales tales como las reglas del lenguaje. Desde esta perspectiva, la forma y la intencionalidad, las propiedades que definen al nivel simbólico de la explicación clásica de los procesos cognitivos en el paradigma dominante, serían, en el mejor caso, aproximaciones útiles para definir las propiedades emergentes del funcionamiento de las redes neurales.

Este nuevo enfoque implica una doble operación conceptual: por una parte, la sustitución de la metáfora clásica de Turing, la metáfora del ordenador, por la metáfora del cerebro. En segundo lugar, la formulación de una pregunta que, sin embargo, sólo se ha hecho posible gracias a la metáfora del ordenador: ¿Cómo puede computar un sistema que posee las propiedades del sistema nervioso?; ¿qué tipo de sistema de cómputo es una red neural? Es evidente que esta nueva metáfora, la del cerebro como sistema de cómputo, puede ayudar a disolver esa ruptura lógica de que habla Tindall; puede proporcionarnos el «órgano intelectual», que él añora, para pasar de «un pensamiento definido» a «una acción molecular definida». En suma, parece que la perspectiva conexionista ofrece la promesa de plantear en términos más tratables el viejo desafío dualista. ¿Es así realmente?; ¿hasta qué punto se justifican las pretensiones de los investigadores conexionistas?

Vimos, en el capítulo cuarto, que los modelos de cómputo de los conexionistas poseen propiedades que resuelven algunos de los aspectos débiles de las sistemas clásicos de procesamiento serial: en primer lugar, son capaces de tener en cuenta restricciones simultáneas y múltiples, tales como las que se dan frecuencia en los problemas que los organismos tienen que resolver con relación al medio. Lo que hacen, para ello, es encontrar las «soluciones óptimas» de las fuerzas de conexión de las redes neurales; es decir, aquellas soluciones que satisfacen en el mayor grado el conjunto de restricciones que se dan en una situación. Además, pueden enfrentarse con éxito a condiciones estimulares degradadas, o en que no se ofrece información acerca de todos los parámetros necesarios para alcanzar situaciones, y éstas son las condiciones en que suelen desenvolverse los organismos. En tercer lugar, los modelos de procesamiento distribuido paralelo muestran un «patrón de degradación elegante»: a diferencia de las máquinas simbólicas de los teóricos clásicos, pueden seguir procesando relativamente bien aunque se dañen conjuntos amplios de unidades. Como sucede con las lesiones cerebrales reales, el deterioro de su actuación se relaciona globalmente con el número de unidades neurales afectadas. Finalmente, los modelos conexionistas son mucho más realistas que los clásicos en lo que se refiere al tiempo de cómputo: las neuronas son considerablemente más lentas que los componentes electrónicos de los ordenadores digitales actuales. Sólo son capaces de realizar alrededor de cien operaciones en el tiempo empleado para solucionar problemas de cierta complejidad computacional. Esta restricción del «programa de los 100 pasos» (Feldman, 1985) lleva a supo-

UNIVERSIDAD DE ANTIOQUIA
BIBLIOTECA CENTRAL

ner que las actividades de los sistemas de cómputo de los organismos deben realizarse en paralelo, pues, en caso contrario, no sería posible explicar la realización de tareas complejas de cómputo en intervalos de tiempo breves.

Para explicar, en líneas generales, el funcionamiento de un sistema conexionista, podemos poner el ejemplo de una máquina a la que queremos hacer capaz de reconocer los fonemas de un lenguaje natural. La solución de esta tarea es mucho más compleja de lo que parece a primera vista: la investigación psicolingüística ha demostrado, hace tiempo, que no existe una relación puntual entre los parámetros físicos de los sonidos del lenguaje y su definición como fonemas. No hay una lista simple de condiciones suficientes y necesarias para reconocer un fonema como una a, por ejemplo. La interpretación de los fonemas es muy sensible a condiciones del contexto, e incluye sonidos muy diferentes dentro de las mismas categorías fonéticas, implica la «percepción categórica» de sonidos que, desde un punto de vista físico, se sitúan en un continuo. Como dice Churchland (1990), «ser una a parece ser cuestión de estar suficientemente cerca de una **a típica** en un número **suficiente** de diferentes **dimensiones** de **relevancia»** (p. 204).

Supongamos entonces que establecemos, en primer lugar, una capa de unidades-neuronas de input, y otra de out-put. La primera estaría compuesta por unidades que modifican sus estados de activación (y consiguientemente sus fuerzas de conexión) al transducir los cambios físicos en los estímulos sonoros del medio, dentro de la gama de sonidos en los umbrales de frecuencia y longitud de onda que caracterizan al habla humana. La segunda capa, la de las unidades de salida, estaría compuesta de un número mucho menor de unidades: una para cada fonema discriminable en el lenguaje natural con el que va a tratar la red neural. Así, podemos imaginar que las unidades de entrada cuentan con algo semejante a un espectrograma de sonidos de habla y las de salida tienen la misión de «categorizar» adecuadamente las dimensiones del espectro en fonemas del lenguaje natural.

Para realizar ese proceso de categorización será necesario incluir en el sistema una o más capas de unidades «ocultas», entre las de entrada y las de salida. En función de los niveles de activación, las fuerzas de conexión y los umbrales de in-put y out-put de las unidades, cuando las unidades de entrada modifican sus estados de activación, extienden la activación a las unidades ocultas y éstas a las de out-put. En principio, el sistema funcionará de forma aleatoria: si establecemos arbitrariamente

las fuerzas de conexión (es decir, lo que se llama el «patrón de conectividad») en el estado inicial del sistema, éste tendrá un funcionamiento completamente azaroso y, en un primer momento, categorizará incorrectamente los sonidos como fonemas. Pero supongamos que dotamos al sistema de dos componentes más: uno externo y otro interno. El externo sería un «profesor». Lo que hace este componente es definir los vectores de salida del sistema (es decir, las listas de niveles de activación de las unidades y fuerzas de conexiones entre ellas), y su distancia con relación a los que hubieran sido los vectores «correctos». El componente interno sería un principio de aprendizaje, inherente al sistema, como por ejemplo la llamada «regla delta generalizada» (Rumelhart, Hinton y William, 1986), que permite calcular los cambios en el vector de out-put (que, a su vez, depende del sistema como un todo) que minimizan la distancia con relación al modelo enseñado por el «profesor». Gracias a las conexiones descendentes, estos cambios se propagarán «hacia atrás» (o «hacia abajo») por todo el sistema. Si se repiten ensayos suficientes, la regla delta generalizada (o algún otro de los principios de aprendizaje de las redes neurales propuestos por los conexionistas, como por ejemplo las elaboraciones del principio de Hebb) garantiza que finalmente el sistema pueda categorizar correctamente los fonemas.

Con modelos de este tipo se han resuelto tareas que resultan extremadamente difíciles para los procesadores seriales clásicos. Por ejemplo, Rosenberg y Sejnowski (1987) han desarrollado una red, a la que llaman **NETtalk**, que es capaz de tomar, como entradas, segmentos de siete letras, y ofrecer, como salidas, codificaciones vectoriales de fonemas. Estas salidas pueden, luego, convertirse en sonidos audibles mediante un sintetizador de sonidos. La red aprende, por tanto, a transformar palabras escritas en habla. Al principio, produce un balbuceo ininteligible. Después de algunos ensayos de aprendizaje (varios miles), y en función de un algoritmo de reducción paulatina de las distancias entre vectores de salida y vectores correctos, alcanza un patrón de conectividad que reduce prácticamente a cero esas distancias: por entonces, su «habla» es perfectamente inteligible, y sólo contiene errores infrecuentes y de menor cuantía.

Un aspecto fundamental de los modelos conexionistas es su dependencia de procedimientos matemáticos que, mediante sistemas de tratamiento y cálculo de vectores y matrices y en función de «algoritmos de aprendizaje», permiten alcanzar soluciones óptimas al problema de cómo conjugar simultáneamente restricciones múltiples. El funcionamiento de

estos sistemas es global y estadístico. Pueden entenderse como sistemas muy complejos que tienden, de forma inherente, a estados de equilibrio. Sin embargo, plantean también un problema importante, que han visto con perspicacia Fodor y Pylyshyn (1988): ¿Hasta qué punto pueden llegar a imitar procesos de manipulación simbólica mediante reglas sensibles a la estructura, tales como los que parecen darse en el lenguaje?; ¿hasta qué punto pueden, incluso, llegar a abstraer propiamente «estructuras»? Parece que su estilo global y estadístico de funcionamiento, que permite alcanzar soluciones óptimas a estados de hechos en que se dan restricciones simultáneas múltiples, no puede llegar a producir un funcionamiento realmente guiado por reglas y estructuras.

Pero, ¿es que, acaso, se guían realmente por reglas y estructuras formales los procesos cognitivos humanos? Para muchos conexionistas, el lenguaje de las reglas y las estructuras de conocimiento es sólo una forma útil de referirse al funcionamiento de las redes neurales metafóricamente, y desde un nivel molar —que prescinde de las inevitables «impurezas» de un sistema conexionista—. Otros, como Clark (1989) piensan que, finalmente, será necesario reconocer un doble nivel para explicar los procesos cognitivos humanos: un plano de «microcognición», de naturaleza subsimbólica, que correspondería el modo de funcionamiento de los sistemas conexionistas. Otro, de carácter simbólico y obediente a reglas y estructuras, que se situaría por encima del anterior. Sin embargo, esta solución no deja de plantear algunos problemas conceptuales importantes: aparte de su falta de parsimonia, parece entremezclar niveles explicativos distintos y se enfrenta a la dificultad del excesivo poder de cómputo, y la excesiva libertad de construcción teórica que ofrece a los investigadores.

En cualquier caso, la «intuición de la forma», de la que hablábamos en el capítulo noveno, se ofrece como obstáculo importante para los modelos conexionistas. Estos no dejan de ser elaboraciones, aunque mucho más complejas de las viejas visiones asociacionistas de la mente: aquellas visiones ligadas a la contingencia, y que nunca pudieron explicar la intuición de necesidad y de forma que se produce frecuentemente en relación con los procesos de razonamiento y lenguaje. Por otra parte, las redes neurales desarrollan propiedades de funcionamiento a las que podríamos llamar «esquemas» y «conceptos», que aúnan un alto poder de codificación con una flexibilidad de la que suelen carecer los objetos intencionales de los modelos psicológicos clásicos del conocimiento.

Como ha destacado Pagels (1991), las redes neurales y los modelos

conexionistas forman parte de **Los sueños de la razón** de las dos últimas décadas de nuestro siglo: de los sueños que se arremolinan alrededor de conceptos como los de orden, complejidad y caos, y que han llevado al desarrollo interdisciplinar de las que empiezan a llamarse «ciencias de la complejidad». El origen de estas ciencias reside, en buena medida, en el estudio de sistemas definidos por enjambres de dimensiones y cuya dinámica produce extraños fenómenos, tales como las bifurcaciones (amplificaciones de pequeñas variaciones iniciales), las caídas en «cuencas de atracción», los ciclos límite, etc. Los fenómenos metereológicos, por ejemplo, corresponden al funcionamiento de sistemas complejos. Así, la mente se acerca cada vez más a la metáfora de la nube, y cuestiona la del reloj: se asemeja más al aspecto pluriforme y difuso de los sistemas complejos que al perfil neto de los sistemas deterministas clásicos. El empleo de nuevas metáforas de la mente, más coherentes con su complejidad real, está permitiendo a los psicólogos acercarse gradualmente a su objetivo final, lejano y desmesurado, de conocer la mente.

BIBLIOGRAFIA

ANDERSON, J. R. (1976) *Language, Memory and Thouhgt.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

ANDERSON, J. R. (1978) «Arguments concerning representations for mental imagery». *Psychological Review*, 85, 249-277 (traducción española: «Argumentos acerca de las representaciones mediante la capacidad para formar imágenes mentales». En SEBASTIAN, M. V. (Ed.) *Lecturas de Psicología de la Memoria.* Madrid: Alianza Universidad, 1983).

ANDERSON, J. R. (1979) «Further arguments concerning representations for mental imagery: A response to Hayes-Roth and Pylyshyn». *Psychological Review*, 86, 249-277.

ANDERSON, J. R. (1983a) *The Architecture of Cognition.* Cambridge, Mass.: Harvard University Press.

ANGUERA, M. T. y VEA, J. J. (Eds.) *Conducta animal y representaciones mentales.* Barcelona: PPU.

ARACIL, J. (1986) *Máquinas, Sistemas y Modelos.* Madrid: Tecnos.

ARBIB, M. A. (1964) *Brains, Machines and Mathematics.* Nueva York:McGraw-Hill (Traducción española: *Cerebros, Máquinas y Matemáticas.* Madrid: Alianza Editorial, 1987).

ASENSIO, M. (1990) *Proyecto docente e investigador.* Madrid: UAM.

ASTINGTON, J. V., HARRIS, P. L. y OLSON, D. R. (Eds.) (1988) *Developing Theories of Mind.* Cambridge: Cambridge University Press.

BAARS, B. J. (1988) *A Cognitive Theory of Consciousness.* Cambridge: Cambridge University Press.

BACHELARD, G. (1948) *La Formación del Espíritu Científico.* Madrid: siglo XXI.

BARTLETT, F. C. (1932) *Remembering: A Study in Experimental and Social Psychology.* Cambridge: Cambridge University Press (Traducción española: *Pensamiento: un Estudio de Psicología Experimental y Social.* Madrid: Debate, 1988).

BENNET, J. (1976) *Linguistic Behaviour.* Cambridge: Cambridge University Press.

BLUMENTHAL, A. L. (1970) *Language and Psychology. Historical Aspects of Psycholinguistics.* Nueva York: Wiley.

BLUMENTHAL, A. L. (1975) «A Reappraisal of Wilhelm Wundt». *American Psychologist,* págs. 1081-1088.

BODEN, M. (1990) «Escaping fron the chinese room». En BODEN, M. (Ed.) *The Psychology of Artificial Intelligence.* Oxford: Oxford University Press, 1990, págs. 89-104.

BOLTER, D. J. (1984) *Turing's Man. Western Culture in the Computer Age.* Chapel Hill: The University of North Carolina Press. (Traducción española: *El Hombre de Turing.* Méjico: Fondea de Cultura Económica, 1988).

BOLTZMANN, L. (1986) *Escritos de Mecánica y Termodinámica.* Madrid: Alianza Editorial.

BOOLE, D. J. (1854) *An Investigation of the Laws of Thought, on which are founded the Matematical Theories of Logic and Probabilities.* Londres.

BORING, E. G. (1929) *A History of Experimental Psychology* (Traducción española de la *2.'* edición de 1950: *Historia de la Psicología Experimental.* Méjico: Trillas, 1988).

BRACEWELL, R. J. Y HIDI, S. E. (1974) «The solution of an inferential problem as a function of the stimulus materials». *Quarterly Journal of Experimental Psychology,* 26, págs. 480-488.

BRANSFORD, J. D. y JONHSON, M. K. (1973) «Considerations on some problems in comprehension». En CHASE, W. G. (Ed.) *Visual Information Processing.* Nueva York: Academic Press.

BRENTANO, F. (1874) *Psychologie von empirischen Standpunkt.* Leipzig (ed. en Hamburgo: Meiner, 1955).

BRETHERTON, I. y BEEGHLY, M. (1982) «Talking about internal states: The acquisition of an explicit theoty of Mind». *Developmental Psychology,* 18 (6), págs. 906-921.

BRETHERTON, I., FTRIZ, J., ZAHN-WAXLER, C. y RIDGEWAY, D. (1986) «Learning to talk about emotions: A funcionalist perspective». *Child Development.* 57 (3), págs. 529-548.

BRETON, PH. (1989) *Historia y Crítica de la Informática.* Madrid: Cátedra, colección Teorema.

BROWN, C., KEATS, J. A., KEATS, D. M. y SEGGIE, I. (1980) «Reasoning about implication: A comparison of malaysian and australian subjeets». *Journal of(Gross-Gultural'Psyékólogy,* 11'1,'págs. B95-410.

BRUNER, J. (1974) *Beyond the Information Given.* Londres: Unwin.

BRUNER, J. (1984) *Acción, Pensamiento y Lenguaje.* Madrid: Alianza psicología.

BRUNER, J. (1986) *Actual Minds, Possible Worlds.* Cambridge: Cambridge University Press. (Traducción española: *Realidad Mental y Mundos Posibles.* Barcelona: Gedisa, 1988).

BRUNER, J., GOODNOW, J. J. y AUSTIN, G. A. (1956) *A Study of Thinking.* Nueva York: Wiley. (Traducción española: *El Proceso Mental en el Aprendizaje.* Madrid: Narcea, 1984).

BRUNSWICK, E. (1937) «Psychology as a science of objective relations». *Phil. Science,* 4, págs. 227-260.

CAPARROS, A. y CABUCIO, F. (1986) La aparición del paradigma cognitivo: una cuestión problemática. *Revista de Historia de la Psicología,* '7 (2), págs. 53-58.

CARELLO, C., TURVEY, M. T., KUGLER, P. N. y SHAW, R. E. (1982) «Inadequacies of the computer metaphor». En GAZZANIGA, M. S. (Ed.) *Handbook of Cognitive Neuroscience.* Nueva York: Plenum

CARNAP, R. (1954) «On Belief Sentences. Reply to Alonso Church». En MAC DONALD, M. (Ed.) *Philosophy and Analysis.* Oxford: Basic Blackwell. págs. 128-131. (Traducción española: «Sobre las oraciones de creencia (respuesta a Alonso Church)». En SIMPSON, T. M. (Ed.) *Semántica Filosófica: Problemas y Discusiones.* Buenos Aires: Siglo XXI, págs. 331-334).

CARRETERO, M. y GARCIA MADRUGA, J. (1984a) «Psicología del pensamiento: Aspectos históricos y metodológicos». En CARRETERO, M. Y GARCIA MADRUGA, J. (Eds.) *Lecturas de Psicología del Pensamiento.* Madrid: Alianza Psicología, págs. 19-42.

CASSIRER, E. (1964) *Philosophie der Symbolischen Formen. Zweiter Teil. Das Mythischen Denken.* Darmstadt: Wissenschafthische Buchgesellschaft. (Traducción española: *Filosofía de las Formas Simbólicas.* México: Fondo de Cultura Económica, 1972).

CLARK, A. (1989) *Microcognition. Philosophy, Cognitive Science and Parallel Distributed Processing.* Cambridge, Mass.: The MIT Press. A Bradford Book.

CLARK, H. H. (1965) «Some structural properties of simple active and passive sentences». *Journal of Verbal Learning and Verbal Behavior,* vol. 4, págs. 365-370.

COHEN, G. (1977) *The Psychology of Cognition.* Londres: Academic Press (Traducción española: *Psicología Cognitiva.* Madrid: Alhambra, 1983).

COLINS, A. (1977) «Why Cognitive Science?» *Cognitive Science,* 1, págs. 3-4.

COLOM, R. y JUAN ESPINOSA, M. de (1990) *Estudios sobre los Fundamentos de la Cognición.* Valencia: Promolibro.

COOPER, L. A. Y SHEPARD, R N. (1973) «Chronometric studies of the rotation of mental images». En CHASE, W. G. (Ed.) *Visual Information Processing.* Nueva York: Academic Press.

COSMIDES, L. (1989) «The Logic of Social Change: Has Natural Selection shaped how humans reason? Studies with the Wason Selection task». *Cognition,* 31, 187-176.

COX, J. R. y GRIGGS, R. A. (1982) «The effeets of experience on performance in Wason's selection task». *Memory and Cognition,* 10, págs. 496-502.

CHENG, P. y HOLYOAK, K. (1985) «Pragmatic Reasoning Schemas». *Cognitive Psychology,* 17, págs. 391-416.

CHISHOLM, R. (1975) «Sentences about believing». *Minnesota Studies in the Philosophy of Science,* vol. 2. Minneapolis: University of Minnesota Press, págs. 510-520. (Traducción española: «Oraciones de creencia». En SIMPSON, T. M. (Ed.) *Semántica Filosófica: Problemas y Discusiones.* Buenos Aires: Siglo XXI, 1973, págs. 417-437).

CHOMSKY, N. (1957) *Syntactic Structures.* La Haya: Mouton. (Traducción española: *Estructuras Sintácticas,.* Madrid: Siglo XXI, 1974).

CHOMSKY, N. (1965) *Aspects of a Theory of Syntax.* Cambridge, Mass.: The MIT Press. (Taducción española: *Aspectos de la Teoría de la Sintaxis.* Madrid: Aguilar, 1970).

CHOMSKY, N. (1980) *Rules and Representations.* Nueva York: Columbia University Press. (Traducción española: *Reglas y Representaciones.* Méjico: Fondo de Cultura Económica, 1983).

CHOMSKY, N. (1981) *Reflexiones acerca del lenguaje.* México: Trillas.

CHOMSKY, N. (1985) *Knowledge of Language: Its Nature, Origins and Use.* Nueva york: Praeger. (Traducción española: *El Conocimiento del Lenguaje. Su Naturaleza, Origen y Uso.* Madrid: Alianza Editorial, 1989).

CHURCH, A. (1954) «Intentional Isomorphism ans Identity of belief». *Philosophical Studies,* 5 (5), págs. 65-71. (Traducción española: «Isomorfismo intencional e identidad de creencia». En SIMPSON, T. M. (Ed.) *Semántica Filosófica: Problemas y Discusiones.* Buenos Aires: Siglo XXI, 1973, págs. 171-182).

CHURCH, A. (1956) «Propositions and sentences». En SIMPSON, T. M. (Ed.) *The Problem of Universals.* Notre Dame, Indiana: University of Notre Dame Press, págs. 3-11. (Tradución española: «Proposiciones y oraciones con una objeción a Scheffer». En SIMPSON, T. M. (Ed.) *Semántica Filosófica: Problemas y Discusiones.* Buenos aires: Siglo XXI, págs. 363-370).

CHURCHLAND, D. M. (1979) *Scientific Realism and the Plasticity of Mind.* Cambridge: Cambridge University Press

CHURCHLAND, P. S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind.* Brain, Cambridge, Mass.: MIT Press.

CHURCHLAND, P. S. (1990) «Cognitive activity in Artificial Neural Networks». En OSHERION, O.N. y Smith, E.E. (Eds) *An Invitation to Cognitive Science Thinking.* Cambridge. MGSS: The MIT Press.

D'ANDRADE, R. G. (1980) «Reasoning and the Wason problem». Manuscrito presentado en el *Symposium on the Ecology of Cognition: Biological, Cultural and Historical Perspectives.* Greensboro' N., California

DANZIGER, K. (1980) «The history of introspecction reconsidered». *Journal of the History of Behavioral Sciences,* 16, págs. 241-262.

DAVIDSON, D. (1965) «Theories of Meaning and Learnable Languages». En BARHILLEL, Y. (Ed.) *The international Congress for Logic, Methodology and Philosophy of Science.* Amsterdam: North Holland, págs. 383-393. (Traducción española: «Teorías del significado y lenguajes aprendibles (una objeción a Quine, Church y Scheffler)». En SIMPSON, T. M. (Ed.) *Semántica Filosófica: Problemas y Discusiones.* Buenos Aires: Siglo XXI, 1973, págs. 383-394).

DAVIDSON, D. (1981) «The material mind». En HAUGEtAD, J. (Ed.) *Mind Design.* Cambridge: The MIT Press. "

DEAÑO, (1974) *Introducción a la Lógica Formal.* Madrid: Alianza.

DEAÑO, A. (1980) *Las Concepciones de la Lógica.* Madrid: Taurus.

DEHN, N. y SCHANK, R. (1982) «Artificial and human intelligence». En STERNBERG, R.J. (Ed.) *Handbook of Human Intelligence.* Cambridge: Cambridge university Press.

DELBRUCK, M. (1986) *Wahrheit und Wirlichkeit - uber die Evolution des Erkennens.* Hamburgo: Rasch und Rohring Verlag. (Traducción española: *Mente y materia. Ensayo de Epistemología Evolutiva.* Madrid: Alianza Editorial, 1989).

DELVAL, J. A. (1975) *El Animismo y el Pensamiento Infantil.* Madrid: Siglo XXI.

DENIS, M,, ENGELKAMP, J. y RICHARDSON, J. (Eds.) (1988) *Cognitive and Neurological Approaches to Mental Imagery.* Dordrecht: Martinus Nijhoff Publishers.

DENNET, D. C. (1978) *Brainstorms. Philosophical Essays on Mind and Psychology.* Cambridge: MIT Press.

DESCARTES, R. (1637) *Discours de la Methode.* (Traducción española: *El Discurso del Método.* Madrid: Akal, 1989.)

DEWART, L. (1989) *Evolution and Consciousness: the Role of Speech in the Origin and Development of Human Nature.* Toronto: University of Toronto Press.

DEWEY, J. (1910/1933) *How We Think.* Lexington, Mass.: D.C. Heath and Comp. (Traducción española *Cómo Pensamos. Nueva Exposición de la Relación entre Pensamiento Reflexivo y Proceso Educativo.* Barcelona: Paidós, 1989).

DRESTSKE, F. (1981) *Knowledge and the Flow of Information.* Cambridge, Mass.: MIT Press.

DRESTSKE, F. (1988) *Explaining Behavior. Reasons in a World of Causes.* Cambridge, Mass.: The MIT Press. A Bradford Book.

DREYFUS, H. (1979) *What Computers can't do.* Nueva York: Harper and Row.

EDWARDS, W. (1968) «Conservatism in human information processing». En KLEINMUNTZ, B. (Ed.) *Formal Representations of Human Judgment.* Nueva York: Willey & Sons.

ELSTEIEN, A. S. y BORDAGE, G. (1984) «Psicología del razonamiento clínico». En M. CARRETERO y J. GARCIA MADRUGA (Eds.) *Lecturas de Psicología del Pensamiento.* Madrid: Alianza psicología, págs. 255-266.

ERICSSON, K. A. y SIMON, H. A. (1984) *Protocol Analysis: Verbal Reports as Data.* Cambridge, Mass: The MIT Press.

EVANS, J. St. B. T. y WASON, P. N. (1976) «Rationalisation in reasoning task». *British Journal of Psychology,* 63, págs. 205-212.

FECHNER, G. T. *Nanna, oder über das Seelenleben der Pflanzen.* (Resumido en LOWRIE, W. (Ed.) *Religión of a Scientist: Selections from Gustav Th. Fechner.* Nueva York: Pantheon Books, 1946)

FECHNER, G. T. (1851) *Zendavesta, oder über die Dinde des Himmel und des Jenseits.*

FELDMAN, J. A. (1985) «Connectionist models and their aplications: introduction». *Cognitiva Science, 9,* págs. 1-2.

FINKE, R. A. (1989) *Principies of Mental Imagery.* Cambridge: The MIT Press.

FLORES, F. y WINOGRAD, T. (1989) *Understanding Computers and Cognition.* Nueva Jersey: Ablex Publishing Corporation (Traducción: española *Hacia la Comprensión de la Información y la Cognición.* Barcelona: Editorial Hispano Europea, 1989)

FODOR, J. (1968) *Psychological Explanation: An Introduction to the Philosophy of Psychology.* Nueva York: Random House. (Traducción española: *La explicación Psicológica.* Madrid: Cátedra, 1980)

FODOR, J. (1975) *The Language of Thought.* Nueva York: Harper & Row. (Traducción española: *El Lenguaje del Pensamiento.* Madrid: Alianza Editorial, 1984).

FODOR, J. (1981) *Representations.* Cambridge, Mass.: The MIT Press.

FODOR, J. (1983) «Fijación de creencias y adquisición de conceptos». En M. PIATTELLI-PALMERINI (Ed.) *Teorías del Lenguaje. Teorías del Aprendizaje.* Barcelona: Crítica, págs. 187-193.

FODOR, J. (1988) *Psychosemantics. The Problem of Meaning in the Philosophy of Mind.* Cambridge, Mass.: The MIT Press. A Bradford Book.

FODOR, J. A. y PYLYSHYN, Z. W. (1988) «Conectionism and cognitive architecture: a critical analysis». *Cognision,* 28. págs. 3-71.

GARCIA ALBEA, J. E. (1986) *Percepción y Computación.* Madrid: Pirámide.

GARDNER, H. (1985) *The Mind's New Science. A history of Cognitive Revolution.* Nueva York: Basic Books. (Traducción española *La Nueva Ciencia de la Mente. Historia de la Revolución Cognitiva.* Buenos Aires: Paidós, 1987).

GEERTZ, C. (1973) *The Interpretation of Cultures.* Nueva York: Basic Books. (Traducción española *La Interpretación de las Culturas.* Barcelona: Gedisa, 1989, tercera reimpresión).

GENTNER, D. y GUDRIN, J. (1985) «The evolution of mental metaphors in Psychology: A 90-year retrospective». *American Psycologist,* febrero, 1985, págs. 181-192.

GIBSON, J. J. (1966) *The Senses Considered as Perceptual Systems.* Boston: Mifflin.

GÖDEL, K. (1931) «Über formal unentscheidbare Sütze der Principia Mathematica un Verwandter Systeme, I». *Monatshefte" für Mathematica und Phisics,* 173-189.

GOODALL, J. (1986) *En la Senda del Hombre.* Barcelona: Salvat.

GREENE, J. M. (1970) «The semantic function of negatives and pasives». *British Journal of Psychology,* vol. 61, págs. 1722.

GREENE, J. (1972) *Psycholinguistics. Chomsky and Psychology.* Londres: Penguin.

GRIFFIN, D. R. (1986) *El Pensamiento de los Animales.* Barcelona: Ariel.

GRIGGS, R. A. (1983) «The role of problem content in the selection task and the THOG problem». En Evans, J. St. B. T. EVANS (Ed.) *Thinking and Reasoning.* Londres: Routledge & Kegan Paul.

GRIGGS, R. A. Y COX, J. R. (1982) «The elusive thematicmaterials effect in Wason's selection task». *British Journal of Psychology,* 73, págs. 407-420.

GRIGGS, R. A, y COX, J. R. (1983) «The effects of problem content and negation on Wason's selection task». *Quarterly Journal of Experimental Psychology.* 35a, págs. 519-533.

HALLPIKE, C. R. (1981) *The Foundations of Primitive Thougth.* Oxford University Press. (Traducción española: *Fundamentos del Pensamiento Primitivo.* Méjico: Fondo de Cultura Económica, 1986).

HARRE, R. (1979) *Social Being. A Theory for Social Psychology.* Oxford: Basil Blackwell. (Traducción española: *El Ser Social.* Madrid: Alianza Universidad, 1982).

HAWKING, S. W. (1988) *A Brief History of Time. From the Bing Bang to Blak Holes.* Nueva York: Bantam Books. (Traducción española: *Historia del Tiempo. Del Bing Bang a los Agujeros Negros.* Madrid, Alianza Editorial, 1990).

HAYES-ROTH, F. (1979) «Distinguishing theories of representation. A critique of Anderson's "Arguments concerning mental imagery"». *Psychological Review,* 86, págs. 376-382.

HEBB, D. O. (1949) *The Organization of Behavior.* Nueva York: Wiley. (Traducción española: *La Organización de la Conducta.* Madrid: Debate, 1985)

HENLE, M. (1962) «On the relation between logic and thinking». *Psychological Review,* 69, págs. 366-378.

HENLE, P. (1972) *Language, Thought and Culture.* An Arbor: The University of Chicago Press.

HINTON, G. E., McLELLAND, J. L. y RUMELHART, D. E. (1986) «Distributed representations». En RUMELHART, D. E., McCLELLAND, J. L. Y Grupo de investigación PDP *Parallel Distributed Processing. Explorations in the Microestructure of Cognition, Vol. 1: Foundations.* Cambridge, Mass.; The MIT Press. A Bradford Book, págs. 77-109.

HINTON, G. E. y SEJNOWSKI, T. J. (1986) «Learning and relearning in Boltzmann Machines». En RUMELHART, D. E., McLELLAND, J. L. Y Grupo de investigación PPD. *Parallel Distributed Procesing. Explorations in the Microestructure of Cognition. Vol 1. Foundations.* Cambridge, Mass: The MIT Press. A Bradford Book.

HOFSTADTER, D. F. (1979) *Gódel, Escher, Bach; An Eternal Golden Braid,* Nueva york: Basi Books. (Traducción española: *Gódel, Escher, Bach: un Eterno y Grácil Bucle.* Barcelona: Tusquets Editores, 1987).

HUMPHEY, N. (1983) *Consciousness Regained. Chapters in the Development of Mind.* Oxford: Oxford University Press. (Traducción española: *La Reconquista de la Conciencia. Desarrollo de la Mente Humana.* Méjico: Fondo de Cultura Económica, 1987).

HUMPHREY, G. (1973) *Psicología del Pensamiento. Teorías e Investigaciones.* Méjico: Trillas.

HUTTENLOCHER, J. (1968) «Constructing spatial images: A strategy in reasoning». *Psychological Review,* 75, págs. 550-560.

¡JAOKENDOFE, ¡R. ¡(1987) *'Consciousness '.andUhe 'Computational Mind.* 'Cambridge, Mass.: The MIT Press. A Bradford Book.

JAMES, W. (1890) *The Principies of Psychology.* Nueva York: Holt. (Tercera edición de Cambridge Mass.: Harvard University Press, 1983. Traducción española: *Principios de Psicología.* Méjico: Fondo de Cultura Económica, 1990).

JANIS, J. L. Y FRICK, F. (1943) «The relationships between attitudes toward conclusions and errors in judging local validity of syllogism». *Journal of Experimental Psychology,* 33,'págs. 73-77.

JOHNSON, M. G. (1967) «Syntactic position and rated meaning». *Journal of Verbal Learning and Verbal Behavoir,* vol. 6, págs. 240-246.

JOHNSON, C. N. Y WELLMAN, H. M. (1980) «Children's developing unders-

tanding of mental verbs: remember, know and guess». *Child.Development,* 51, págs. 1095-1102.

JOHNSON-LAIRD P. N. (1968) «The choice of the pasive voice in a communicative task'». *British Journal of Psychology*, vol. 59, págs. 7-15

JOHNSON-LAIRD, P. N. (1975) Models of deduction. En FALMAGNE, R. (Ed.) *Reasoning: Representation and Process in Children and Adults.* Hillsdale, N.J.: Lawrence erlbaum.

JOHNSON-LAIRD, P. N. (1983) *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness.* Cambridge: Cambridge University Press.

JOHNSON-LAIRD, P. N. (1988) *The Computer and the Mind: An introduction to Cognitive Science.* Glasgow: William Collins Sons and Co. Ltd (traducción española de Alfonso Medina: *El Ordenador y la Mente. Introducción a la Ciencia Cognitiva.* Barcelona: Paidós, (1990).

JOHNSON-LAIRD, P. N. (1989) «Mental models»». En POSNER, M. I. (Ed.) *Foundations of Cognitive Science.* Cambridge, Mass: The MIT Press.

JOHNSON-LAIRD, P. N., LEGRENZI, P. y LEGRENZI, M. (1972) «Reasoning and a sense of reality». *British Journal of Psychology,* 63, págs. 395-400

JOHNSON-LAIRD, P. N., LEGRENZI, P. y SONINO LEGRENZI, M. (1977) «Reasoning an a sense of reality». *British Journal of Psychology,* 63, págs. 295-400.

JUAN ESPINOSA, M. de y COLOM, R. (1989) *Psicología Diferencial y Cognición.* Valencia: Promolibro

KINTSCH, W. (1974) *The Representation of Meaning in Memory.* Hillsdale, N. J.: L.E.A.

KINTSCH, W., MILLER, J. R. Y POLSON, R. E. (1984) *Method and Tactics in Cognitive Science.* Hillsdale, N.J.: LEA

KÖHLER, N. (1972) *Psicología de la Forma: su Tarea y Ultimas Experiencias.* Madrid, Biblioteca Nueva.

KOSSLYN, S. M. (1980) *Image and Mind.* Cambridge, Mass.: Harvard University Press.

KOSSLYN, S. M. y POMERANZ, J. R. (1977) «Imagery, propositions and the form of internal representations». *Cognitive Psychology,* 9, 52-76.

KOSSLYN, S. M. y POMERANZ, J. R. (1986) «Imágenes, proposiciones y la forma de las representaciones internas». En GARCIA-ALBEA, J. E. Ed. *Percepción y Computación.* Madrid: Pirámide.

KOSSLYN, S. M. y SCHWARTZ, S. P. (1977) «A data-driven simulation of visual imagery». *Cognitive Science,* 9, 5276.

KUHN, D., PENNINGTON, N. y LEADBEATER, B. (1984) «El pensamiento adulto desde una perspectiva evolutiva: el razonamiento de los jurados». En CARRETERO, M. y GARCIA-MADRUGA, J. (Eds.): *Lecturas de Psicología del Pensamiento,* Madrid: Alianza Universidad, págs. 267-296.

KUHN, T. S. (1977) «Second thought on paradigms». En SUPPE, F. (Ed.) *The Structure of Scientifiec Theories.* Illinois: University of Illinois Press.

LACHMAN, P., LACHMAN, J. L. y BUTTERFIELD, E. C. (1979) *Cognitive Psychology and Information Processes. An introduction.* N. J.: Lawrence Erlbaum.

LAFUENTE, E. (1986) «La significación de Tolman». *Revista de Historia de la Psicología,* 7 (3), págs. 15-30.

LAKATOS, I. (1984) «Falsification and the methodology of scientific research programes». En LAKATOS, I. y MUSGRAVE (Eds.) *Criticism and the Growth of Knowledge.* Cambridge University Press.

LEARY, D. E. (1979) «Wundt and after: Psychologist's shifting relations with the Natural Sciences, Social Sciences and Philosophy». *Journal of History of Behavioral Sciences,* 15, págs. 231-241.

LEARY, D. E. (1990) *Metaphors in the History of Psychology.* Cambridge: Cambridge University Press.

LEAHEY, T. (1981) «The revolution never appened :information processing *is* behaviorism». Comunicación presentada en la 52 reunión anual de la Eastern Psychological Association. Nueva York.

LEFFORD, A. (1946) «The influence of emotional subject matter on logical reasoning». *Journal of General Psychology,* 30, págs. 127-151.

LEIBNIZ, G.G. (1720) *Lehrsatze iiber die Monadologie* (Traducción española de Julián Valverde: *Monadología.* Oviedo: Pentalfa, 1981).

LESLIE, A. M. (1987) «Pretense and representation: the origins of "Theory of mind"». *Psychologocal Review,* 94, págs. 412-426.

LESLIE, A. M. (1988) «Some implications ef pretense for understanding the child's Theory of Mind». En ASTINGTON, J. , HARRIS, P. L. y OLSON, D. R. (Eds.) *Developing Theories of Mind.* Cambridge: Cambridge University Press.

LINDSAY, P. H. Y NORMAN, D. A. (1972) *Human Information Processing.* Nueva York: Academic Press. (Traducción española: *Procesamiento de la Información Humana.* Madrid: Tecnos, 1983).

LOPEZ CEREZO, J. A. (1989) «El caso contra la psicología popular». *Cognitiva,* 2, 3, págs. 227-241.

LORENTE DE NO, R. (1938) «Analysis of the activity of the chains of internuncial neurons». *Journal or Neurophysiology-,* 1, 207-244.

LORENTE DE NO, R. (1939) «Transmission of impulses through cranial motor nuclei». *Journal of Neurophysiology,* 2, 402-464.

LORENTE DE NO, R. (1943) «Cerebral Cortex: Architecture». En FULTON, J. R. (Ed.) *Physiology of the Nervous System.* 2ª ed. Oxford University Press.

MALT, B. C. y SMITH, E. E. (1984) «Correlated Properties in Natural Categories». *Journal of Verbal Learning and Verbal Behavior,* 23, págs. 250-269.

MANKTELOW, K. I. y EVANS, J. SIB. T. (1979) «Facilitation of reasoning by realism: Effect or non-effect?» *British Journal of Psychology,* 70, 477-488.

MARR, D. (1982) *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information.* Nueva York: Freeman and Co. (Traducción española: *La visión. Una Investigación Basada en el Cálculo acerca de la Representación y el Procesamiento Humano de la Información Visual.* Madrid: Alianza, 1985).

McCAULEY, R. N. (1987) Thé Role of Cognitive Explanations in Psychology. *Behaviorism,* 15 (1), págs. 27-40.

McCLELLAND, J. L., RUMELHART, D. E. y Grupo de Investigación PDP (1986) *Parallel Distributed Processing. Explorations in the Microestructure of*

Cognition. Vol. 2: Psychological and Biological Models. Cambridge, Mass.: The MIT Press. A Bradford Book.

McCULLOCH, W. y PITTS, W. (1943) «A logical calculus of the ¡deas inmanent ¡n nervous activity». *Bulletin for Mathematical Biophysics,* 5, 115-155 (reimpreso en BODEN, M. (Ed.) *The Philosophy of Artificial Intelligence,* Oxford: Oxford University Press, 1990, 22-39).

MEHLER, J. (1963) «Some effects of grammatical transformations om the recall of English sentences». *Journal of Verbal Learning and Verbal Behavior,* 2, págs. 346-351.

MERVIS, C. B. y ROSCH, E. (1981) «Categorization of Natural Objects». *Annual Review of Psychology,* 32, 496-522.

MILLER, G. A., GALANTER, E. y PRIBRAM, K. H. (1960) *Plans and the Structure of Bahavior* (traducción española de Rodolfo Fernandez González: *Planes y Estructura de la Conducta.* Madrid: Debate, 1983).

MILLER, G. A. y McKEAN, K. E. (1964) «A chronometic study of some relations between sentences». *Quarterly Journal of Experimental Psychology,* 16, págs. 297-308.

MINSKY, M. (1975) «A framework for representing knowledge». En WINSTON, P. H. (Ed.) *The Psychology of Computer Vision.* Nueva York: McGraw Hill.

MINSKY, M. L. (1986) *La Sociedad de la Mente.* Buenos Aires:. Galápago.

MINSKY, M. L. y PAPERT, S. (1969) *Perceptrons.* Cambridge, Mass.: The MIT Press.

MIRALLES, J. L. (1986) «Lenguaje y cognición en Wundt». *Revista de Historia de la Psicología,* 7 (2), 59-67.

MISCIONE, J. L., MARVIN, R.S., O'BRIEN, R. G. y GREENBERG, M. T.. (1978) «A developmental study of preschool children's understanding of the words "know" and "guess"». *Child Development,* 49, págs. 1107-1113.

MORGAN J. J. y MORTON, J. T. (1944) «The distortion of syllogistic reasoning produced by personal convictions». *Journal of Social Psychology,* págs. 30-59. (Traducción española: «La distorsión del razonamiento silogístico a causa de las convicciones personales». En DELVAL, J. (Ed.) *Lógica y psicología.* Madrid: Alianza Universidad, 1977).

NEISSER, V. (1976) *Cognition and Reality.* San Francisco: W. H. Freeman.

NEWELL, A., SHAW, J. C. y SIMON, H. A. (1958) «Elements of a theory of human problem solving». *Psychological Review,* 65, 151-166.

NEWELL, A. y SIMON, H. A. (1972) *Human Problem Solving.* Englewood Cliffs, N. J.: Prentice-Hall.

NEWELL, A. y SIMON, H. A. (1976) «Computer science as empirical enquiry: Symbols and search. Tenth Turing Lecture». En *Communications of the Association for Computing Machinery,* 19. Association for Computing Ma•chinery (reeditado en BODEN, M. (Ed.) *The Philosophy of Artificial Intelligence.* Oxford: Oxford University Press, 1990, 105132).

NIDDITCH, P. H. (1980) *The Development of Mathematical Logic.* Londres: Routledge (Traducción española: *El desarrollo de la Lógica Matemática.* Madrid: Cátedra, 1980).

NORMAN, D. (1980) «Twelve Issues for Cognitive Science». *Cognitive Science,* 4, 1-32.

OLSON, D. R. y TORRANCE, N. (1987) «Language, literacy and Mental States». *Discourse and Processes, 10,* págs. 157-167.

ONIANS, R. B. (1954) *The Origins of Modern European Thought about the Body, the Mind, the Soul, the World, Time and Fole: New Interpretations of Greek, Román and Kindred Evidence, also of some Basic Jewish and Christian Belief.* Cambridge, Mass: Cambridge University Press.

OSGOOD, C. E., SUICI, G. J. y TANNENBAUM, P. H. (1957) *The Measurement of Meaning.* Nueva York: Wiley.

PAGELS, H.R. (1991) *Los Sueños de la razón* Barcelona. Gedisa.

PAIVIO, A. (1971) *Imagery and Verbal Processes.* Nueva York: Holt, Rinehart and Winston.

PAIVIO, A. (1977) «Images, Propositions and Knowledge». En NICHOLAS, J. M. (Ed.) *Images, Perception and Knowledge,* Dordrech: Reidel Publishing Comp., págs. 47-71.

PAIVIO, A. (1986) *Mental Representations: a Dual Coding Approach.* Nueva York: Oxford University Press.

PAIVIO, A. y BEGG, I. (1981) *Psychology of Language.* New Jersey: Prentice Hall.

PATTEE, H. H. (1977) «Dynamic and linguistic modes of complex systems». *International Journal of General Systems,* 3, págs. 259-266.

PATTEE, H. H. (1982) «The need of complementarity in models of cognitive behavior. A response to Carol Fowler and Michael Turvey». En WEIMER, W. y PALERMO, D. (Eds.) *Cognition and the Symbolic Processes.* vol. II. Hilisdale, N.J.: Laewrence Erlbaum.

PATTEE, H. H. (1984) «Discrete and continuous processes in computers and brains». En CONRAD, M., GUTTINGE, W., y DALCIN, M. (Eds.) *Physics and Mathematics of Nervous System.* Nueva York: Springer-Verlag, págs. 128-148.

PEREZ GOMEZ, A. (1985) «Conocimiento académico y aprendizaje significativo». En GIMENO SACTISTAN, J. y PEREZ GOMEZ, A. (Eds.) *La Enseñanza: su Teoría y su Práctica.* Madrid, Akal, 1985.

PETERSON, C. R. y BEACH, L. R. (1967) «Man as an intuitive statician». *Psychological Bulletin,* 68, págs. 29-46.

PIAGET, J. (1926) *La Representation du Monde chez L'Enfant.* París: Presses Universitaires de France (traducción española: *La Representación del Mundo en el Niño.* Madrid: Morata, 1984, 6a. reed.)

PIAGET, J. (1936) *La Naissance de L'Intelligence chez L'Enfant.* Neuchatel: Delachause et Niestlé. (Traducción española: *El Nacimiento de la Inteligencia en el Niño.* Madrid: Aguilar, 1969).

PIAGET, J. (1947) *La Psychologie de L'Intelligence.* París: A Colin (Traducción española : *La Psicología de la Inteligencia.* Buenos Aires: Psique, 1971).

PIAGET, J. (1949) *Traité de Logique.* París: Colín, 1949.

PIAGET, J. (1953) *Logic and Psychology.* Manchester University Press.

PIAGET, J. (1961) *La Naissance de L'Intelligence chez L'Enfant.* Neuchâtel: Delachaux & Niestlé, 1961 (Traducción española: *El Nacimiento de la Inteligencia en el Niño.* Madrid: Aguilar, 1969).

PIAGET, J. (1970) *Psychologie et Epistémologie.* París: Denoel. (Traducción española: *Psicología y Epistemología.* Barcelona: Ariel, 1971).
PIAGET, J. (1974) *La prisé de Conscience.* París: Presses Universitaires de France (Traducción española: *La toma de Conciencia.* Madrid, Morata, 1976).
PINILLOS, J. L. (1981) «Wundt y la explicación psicológica». *Revista de Historia de la Psicología,* 2 (4), págs. 355360.
POLLARD, P. (1982) «Human reasoning some posible effects of availability». *Cognition,* 12, págs. 65-96.
POSNER, M. I. (eD.) (1989) *Foundations of Cognitive Science.* Cambridge, Mass.: The MIT Press. A Bradford Book.
POZO, J. I. (1987) *Aprendizaje de la Ciencia y Pensamiento Causal.* Madrid, Visor.
PREMACK, D. y PREMACK, A. J. (1983) *The Mind of an Ape.* Nueva York: Norton (Traducción española: *La Mente del Simio.* Madrid: Debate, 1988).
PREMACK, D. Y WOODRUFF, G. (1978) «Does the chimpanzee have a theory of mind?» *The Behavioral and Brain Sciences,* I, 515-526.
PRIGOGINE, I. (1983) *¿Tan sólo una ilusión? Una Exploración del Caos al Orden.* Barcelona: Tusquet, 1983.
PUTNAM, H. (1960) «Minds and Machines». En ANDERSON, A. R. (Ed.) *Minds and Machines.* Nueva York: • Prentice-Hall (Traducción española: «Mentes y máquinas». En ANDERSON, A. R. (Ed.) *Mentes y máquinas.* Barcelona: Tusquets, 1984, págs. 113-149)
PUTNAM, H. (1981) «Reductionism and the Nature of Psychology». En HAUGELAND, J. (Ed.) *Mind Design.* Cambridge, Mass.: The MIT Press.
PYLYSHYN, Z. W. (1973) «What the mind's eye tells the mind's brain: A critique of mental imagery». *Psychological Bulletin,* 80, 1-24.
PYLYSHYN, Z. W. (1978) «Imagery and Artificial Intelligence». En SAVAGE, W. (Ed.) *Perception and Cognition. Issues in the Foundation of Psychology. The Minnesota Studies in the Philosophy of Science.* Vol. IX. Minneapolis: University of Minnesota Press, págs. 15-55.
PYLYSHYN, Z. W. (1979) «Validating computational models: A critique of Anderson's indeterminacy of representation claim». *Psychological Review,* 86, 383-394.
PYLYSHYN, Z. W. (1980) «Computation and cognition: Issues in the foundations of Cognitive Science». *The Behavioral and Brain Sciences,* 3 (1), 111-132.
PYLYSHYN, Z. W. (1981) «The imagery debate: Analogue media versus tacit knowledge». *Psychological Review,* 88, 1645.
PYLYSHYN, Z. W. (1983) «La naturaleza simbólica de las representaciones mentales». M. V. SEBASTIAN (Ed.) *Lecturas de Psicología de la Memoria.* Madrid: Alianza Universidad.
PYLYSHYN, Z. W. (1984) *Computation and Cognition: Toward a Foundation for Cognitive Science.* Cambridge, Mass.: MIT Press. A Bradford Book (Traducción española: *Computación y Conocimiento. Hacia una Fundamentación de la Ciencia Cognitiva.* Madrid: Debate, 1988).
PYLYSHYN, Z. W. (1986a) «Imágenes e inteligencia artificial». En A. GARCIA-ALBEA (Ed.), *Percepción y Computación.* Madrid: Pirámide.
PYLYSHYN, Z. W. (1986b) «Qué le dice el ojo de la mente al cerebro de la mente: crítica de las imágenes mentales». En A. GARCIA-ALBEA (Ed.), *Percepción y Computación.* Madrid: Pirámide.
PYLYSHYN, Z. W. (1989) «Computing in Cognitive Science». En M. I. POSNER (Ed.) *Foundations of Cognitive Science.* Cambridge, Mas.: The MIT Press, págs. 49-92.
QUINE, W. (1960) *Word and Object.* Cambridge: The MIT Press.
READ, K. E. (1955) «Morality and the concept of the person among, the Gahuku-Gama». *Oceania,* 25 (4), págs. 233-282.
RICHARDSON, J. T. (1980) *Mental Imagery and Human Memory.* Londres: The Macmillan Press.
RIVIERE, A. (1985) «Sobre la multiplicidad de las representaciones. Un viaje por los vericuetos de los lenguajes del pensamiento». En MAYOR, J. (Ed.) *Actividad Humana y Procesos cognitivos.* Madrid, Alhambra, págs. 109-129.
RIVIERE, A. (1986) *Razonamiento y Representación.* Madrid: Siglo XXI.
RIVIERE, A. (1987) *El Sujeto de la Psicología Cognitiva.* Madrid, Alianza Editorial.
RIVIERE, A. (1989) «Más a favor de la psicología popular». *Cognitiva,* 2-3, págs. 261-265.
RIVIERE, A. (1990 a) «Mente y conciencia en los «Principios de Psicología»: un diálogo con James cien años después». *Psicothema,* 2 (1), págs. 111-133.
RIVIERE, A. (1990 b) «Procesos pragmáticos y atribución de estados mentales: una análisis de la deficiencias sociales severas en humanos y de las peculiaridades comunicativas en otros primates». *Actas del II Congreso del Colegio Oficial de Psicólogos.* Valencia, 1990.
RIVIERE, A. y CASTELLANOS, J. L. (En preparación) Teoría de la mente y competencias operatorias en autismo.
ROSCH, E. (1978) «Principies of Categorization». En ROSCH, E. y LLOYD, B. B. (Eds.) *Cognition and Categorization.* Hillsdale, N. J.: Lawrence Erlbaum, págs. 27-48.
ROSCH, E. y MERVIS, C. B. (1975) «Family Resemblances: Studies in the Internal Structure of Categories». *Cognitive Psychology, 8,* págs. 382-493.
ROSENBERG, A. (1980) *Sociobiology and the Preemption of Social Science.* Baltimore: John Hopkins University Press.
ROSENBERG, C. R. y SESNOWSKI, T. J. (1987) «Parallel network that learn to pronounce English text». *Complex System,* 1, pp. 145-168.
ROSENBLATT, F. (1959) «Two theorems of statistical separability in the perceptron». En *Mechanization of Thought Processes: Proceedings of a Symposium held at the National Physical Laboratory,* Nov. 1958. Vol. 1., Londres: HM Stationery Office, 421-456.
ROSENBLATT, F. (1962) *Principies of Neurodynamics.* Nueva York: Spartan.
RUMELHART, D. E. (1980) «Schemata: The building blocks of Cognition». En SPIRO, A. *et. al.* (Eds.) *Theoretical Issues in Reading Comprehension.* Nueva Jersey: Laurence Erlbaum.
RUMELHART, D. E. (1982) «Coments on Cognitive Science». Manuscrito sin publicar. San Diego: Universidad de California.
RUMEHLART, D. E. (1984) «Schemata and the cognitive system». En WIER,

R. S. y SRULL, T. K. (Eds.) *Handbook of Social Cognition. Voll.* Nueva York: Laurence Erlbaum, págs. 161-188.

RUMELHART, D. E. (1989) «The architecture of mind». En M. I. POSNER (Ed.) *Foundations of Cognitive Science.* Cambridge, Mas.: The MIT Press, 133-156.

RUMELHART, D. E., McCLELLAND, J. L. y PDP Research Group (Eds.) (1986) *Parallel Distributed Processing Explorations in the Microstructure of Cognition.* Cambridge, Mass: Cambridge University Press.

RUMELHART, D. E. y McCLELLAND, J. L. (1986) «On learning the past tenses of english verbs». En McCLELLAND, J.L. y RUMELHART, D.E. (Eds.) *Parallel Distributed. Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* Cambridge: The MIT Press

RUMELHART, D. E., & ORTONY, A. (1977) «The representation of knowledge in memory». En: Anderson, R. C. et al (eds) *Schooling and the Adquisition of Knowledge.*

RUMELHART, D. E., SMOLENSKY, P., McCLELLAND, J, y HINTON, G. (1986) «Schemata and sequential thought processes in PDP models». En McCLELLAND, D., RUMELHART, D. y PDP Research group, *Parallel Distributed Processing: Explorations in the Microestructure of cognition, vol. 2, págs.* Cambridge Mass,: Cambridge University Press.

RYLE, G. (1949) *The concept of Mind.* Londres: Hutchison.

SAINZ, J. (1985) «Hacia un nuevo modelo de categorización». En MAYOR, J. (Ed.) *Actividad Humana y Procesos Cognitivos (Homenaje a J.L. Pinillos).* Madrid: Alhambra, págs. 81-103.

SAVIN, H. B. y PERCHONOCK, E. (1965) «Grammatical structure and the immediate recall of english sentences». *Journal of Verbal Learning and Verbal Behavior,* vol. 4, págs. 348-353.

SCHANK, R. C. y ABELSON, R. P. (1977) *Scripts, Plans, Goals and Understanding. An Inquiry into Human Knowledge Structures.* Hillsdale, N. J. : LEA. (Traducción española: *Guiones, Planes, Metas y Entendimiento.* Barcelona: Paidós, 1987).

SEARLE, J. (1980) «Minds, brains and programs». *The Behavioral and Brain Sciences,* 3, 473-424 (reeditado en BODEN, M. (Ed.) *The Philosophy of Artificial Intelligence.* Oxford: Oxford University Press)

SEARLE, J. (1983) *Intentionality: An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press.

SEOANE, J. (1984) «Conocimiento y representación social». En MAYOR, J. (Ed.) *Actividad Humana y Procesos Cogitivos (Homenaje a J. L. Pinillos).* Madrid: Alhambra, págs. 81-103.

SHEPARD, R. N. y METZLER, J. (1971) «Mental rotation of three dimensional objects». *Science,* 171, 701-703.

SIMON, H. A. Y ICAPLAN, C. A. (1989) «Foundations of cognitive science». En M. I. POSNER (Ed.) *Foundations of Cognitive Science.* Cambridge, Mas.: The MIT Press, págs. 1-43.

SIMPSON, T. M. (1973) *Semántica Filosófica: Problemas y Discusiones.* Buenos Aires: Siglo XXI.

SIMPSON, T. M. (1975) *Formas Lógicas Realidad y Significado.* Buenos Aires: Editorial Universitaria de Buenos Aires.

SKINNER, J. (1957) *Verbal Behavior.* Nueva Jersey: Pretince Hall, 1957. (Traducción española: *Conducta Verbal,* Méjico: Trillas, 1981).

SLOBIN, D. I. (1966) «Grammatical transformation and sentence comprhension in childhood and adulthood». *Journal of Verbal Learning and Verbal Bahavior,* 5, págs. 219-227.

SMITH, E. E. (1989) «Concepts and induction». En M. I. POSNER (Ed.) *Foundations of Cognitive Science.* Cambridge, Mas.: The MIT Press.

SMITH, E. E. y MEDIN, D. L. (1981) *Categories and Concepts.* Cambridge, Mass.: Harvard Universtity Press.

SMITH, E. E., SHOBEN, E. J. y RIPPS, L. J. (1974) «Structures and processes in semantic memory : A featural model for semantic decisions. *Psychological Review,* 81 (3), 214241.

SOKOLOV, A. N. (1972) *Inner Speech and Thought.* Nueva York: Plenum Press.

SPERBER, D. y WILSON, DT (1986) *Re le vanee: Communication and Cognition.* Oxford: Blackwell.

STICH, (1983) *From Folk Psychology to Cognitive Science.* Cambridge, MA: MIT Press.

TOLMAN, E. C. (1932) *Purpositive Behavior in Animáis and Men.* Nueva York: Appelton Century Crafts.

TREVARTHEN (1977) «Descriptive analysis of infant communicative Behavior». En CHAFFER, H. R. (Ed.) *Studies of Infant-Mother Interaction.* Londres: Academic Press.

TREVARTHEN, C. (1982) «The primary motives for coopetative undeistanding». En BUTTERWORTH, G. y LICHT, P. (Eds.) *Social Conition. Stlidies of the Development of Understanding.* Brighton: Harvester Press.

TURING, A. M. (1936) «On computable numbers, with an application to the Entscheindungproblem». *Proceedings of the London Mathematical Society,* 2ª serie, 42, 230-265 (reimpreso en DAVIS, M. (Ed.) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions.* Hewlett, N.Y.: Raven Press, 1965).

TURING, A. M. (1950) «Computing machinery and intelligence». *Mind,* 59, 433-460.. (Reeditado, en. ANDERSON, A. R. (Ed.) *Controversia sobre Mentes y Máquinas.* Barcelona: Tusquests, 1984, 11-50).

VAN DUYNE, P. C. (1974) «Realism and linguistic complexity in reasoning». *British Journal of Psychology,* 65, págs. 5967.

VAN DUYNE, P. C. (1976) «Necessity and contingency in reasoning». *Acta Psychologica.* 44, 85-101.

VARELA, F. J. (1988) *Cognitive Science. A Cartography of Current Ideas.* (Traducción española: *Conocer. Las Ciencias Cognitivas: Tendencias y Perspectivas. Cartografía de las Ideas Actuales.* Barcelona: Gedisa, 1990)

VEGA, M. de (1981) «Una exploración de los metapostulados de la Psicología contemporánea. El logicismo». *Análisis y Modificación, de Conducta,* 16, 345-376.

VEGA, M. de (1982) «La metáfora del ordenador: implicaciones y límites». En

I. DELCLAUX y J. SEOANE (Eds.) *Psicología Cognitiva y Procesamiento de la Información.* Madrid: Pirámide, págs. 63-81.

VEGA, M. de (1984) *Introducción a la Psicología Cognitiva.* Madrid: Alianza Psicología.

VEGA, M. de (1987) «Re-procesamiento de la información : una breve incursión en los abismos infomacionales». (X) *Boletín de Psicología,* n. 15, junio 1987, Valencia.

VEGA, M. de (1988) «Mental imagery and perception: Modularity or functional equivalences?» En DENNIS.M, ENGELKAMP, J. y RICHARDSON, J. T. (Eds.) *Cognitive and Neuropsychological Approaches to Mental Imagery.* Dordrecht: Martinus Nijhoff Publishers, 47-56.

VEGA, M. de (1989) «Pragmatismos, meta-cogniciones y teorías populares». *Cognitiva,* 2-3, págs. 275-279.

VEGA de Y AGAPITO, M. (1989) *Aproximación al Concepto Modular de la Mente en el Marco del Programa Funcionalista.* Universidad de Salamanca. Facultad de Filosofía y Ciencias de la Educación. Tesina de licenciatura sin publicar.

VOLLMER, G. (1984) «Mesocosm and objective knowledge». En WUKETITS, F. M. (Ed.) *Concepts and Approaches in Evolutionary Epistemology.* Amsterdan: Reidel, págs. 69-121.

VYGOTSKI, L. S. (1979) *El Desarrollo de los Procesos Psicológicos superiores.* Barcelona: Crítica.

WASON, P. C. (1965) «The contexts of plausible denial». *Journal of Verbal Learning and Verbal Behavior,* vol. 4, págs. 7-11

WASON, P. C. (1966) «Reasoning». En B. M. FOSS: *New Horizons in Psychology.* Horchmansworth, Middlesex: Penguin.

WASON, P. C. (1968) Reasoning about a rule. *Quarterly Journal of Experimental Psychology,* 20, págs. 273-281.

WASON, P. C. (1969a) «Regression in reasoning?» *British Journal of Psychology.* 60, 471-480

WASON, P. C. (1978) «Hypothesis testing and reasoning». En: *The Open University: Cognitive Psychology. Unit 25.* Londres: Open University Press

WASON, P. C. (1984) «Realismo y racionalidad en la tarea de selección». En M. CARRETERO y J.A. GARCIA MADRUGA (Eds) *Lecturas de Psicología del Pensamiento.* Madrid:Alianza Psicología, págs. 99-112.

WASON, P. C. y JOHNSON-LAIRD, P. N. (1972) *Psychology of Reasoning.* Londres: Batsford. (Traducción española: *Psicología del Razonamiento.* Madrid: Debate, 1980)

WASON, P. C. & SHAPIRO, D. (1971) «Natural and contrived experience in a reasoning problem». *Quarterly Journal of Experimental Psychology,* 23, págs. 63-71.

WATSON, J. B. (1913) «Psychologie as the behaviorist views it». *Psychological Review.* 20, págs. 158-177.

WATSON, J. B. (1914) *Behavior.* Nueva york: Holt.

WATSON, J. B. (1919) *Psychology from the Standpoint of a Behaviorist.* Filadelfia: Lippincot.

WATSON, J. B. (1924) *Behaviorism.* Chicago: University of Chicago Press.

WATSON, J. B. (1930) *Behaviorism.* Nueva York: Norton (Traducción española: *El Conductismo.* Buenos Aires: Paidós, 1972).

WELLMAN, H. M. y ESTES, D. (1987) «Children early use of mental verbs and what they mean». *Discourse Processes,* págs. 141-156.

WELLMAN. H. M. y JOHNSON, C. N. (1979) «Understanding of mental processes: a developmental study of remember and forget». *Child Development,* 50, 79-89.

WHITEHEAD, A. N. y RUSSELL, B. (1910-1913) *Principia Mathematica.* 3 vols., Cambridge: Cambridge University Press. (2^3 ed. con nueva introducción y apéndice 1925-1927).

WINOGRAD, T. (1981) «¿Qué significa comprender el lenguaje?» En: Norman, D. A. (ed) *Perspectives in Cognitive Science.* New Jersey: LEA (Traducción española) *Perspectivas de la Ciencia Cognitiva.* Barcelona: Paidós, 1987, págs. 275-314.

WITTGENSTEIN, L. (1968) Notas sobre las conferencias sobre «Experiencia privada» y «Datos sensibles». En VILLANUEVA, E. *El Argumento del Lenguaje Privado.* Méjico, Universidad Autónoma de Méjico, 1979).

WHORF, B. L. (1956) *Language, Thought, and Reality.* Cambridge, Mass. MIT Press.

WOODRUFF, G. y PREMACK, D. (1979) «Intentional communication in the chimpanzee: the development of decepción». *Cognition,* 1979, 7, págs. 333-362.

WUNDT, W. (1874) *Grundzüge der physiologischen Psychologie.* Leipzig: Engelmann.

WUNDT, W. (1882) «Die Aufgaben der experimentellen Psychologie». *Unsere Zeit,* (reimpreso en *Essays,* 2a. ed., Leipzig: Engelmann, 1906).

WUNDT, W. (1887) «Selbstbeobatchung und innere Wahrnehmung». *Philosophische Studien,* 4, 292-309.

YACHANIN, S. A. y TWENEY, R. D. (1982) «The effect of thematic content on cognitive strategies in the four-card selection task». *Bulletin of the Psychonomic Society,* 19, 8790.

YELA, M. (1974) *La Estructura de la Conducta. Estímulo, Situación y Conciencia.* Real Academia de Ciencias Morales y Políticas.